Math 775: Homework 1

Alex Dewey

February 22, 2016

1 Exercises

Exercise 1.

(a.)

Prior × Likelihood = Marginal × Posterior.

$$p(D = d)p(LVH = \text{yes}|D = d) = p(LVH = \text{yes})p(D = d|LVH = \text{yes}). \quad (1)$$

We're given both the prior and the likelihood probabilities on the left side of Equation 1, so we can simply multiply them together.

In order to get the marginal probability $p(LVH = \text{yes})$, we simply need to sum over the possible values for D:

$$p(LVH = \text{yes}) = \sum_{i=1}^{6} p(D = d_i)p(LVH = \text{yes}|D = d_i). \quad (2)$$

Then we can simply divide the left side of Equation 1 by the marginal probability $p(LVH = \text{yes})$ to get the posterior proabability for each disease d. Below is a table summarizing the calculations ($p(LVH = \text{yes})$ has been shortened to $p(LVH)$ for space):

Disease	**Prior**	**Likelihood**	**Prior×Likelihood**	**Posterior**
d	$p(D = d)$	$p(\text{LVH}\|D = d)$	$p(D = d)p(\text{LVH}\|D = d)$	$p(D = d\|\text{LVH})$
PFC	0.45	0.10	0.045	**0.3008**
TGA	0.14	0.15	0.021	**0.1404**
Fallot	0.03	0.12	0.036	**0.0241**
PAIVS	0.06	0.90	0.054	**0.3610**
TAPVD	0.12	0.05	0.006	**0.0401**
Lung	0.20	0.10	0.020	**0.1337**
Marginal		$p(\text{LVH})$	**0.1496**	**1**

The posterior distribution $p(D = d|LVH = \text{yes})$ differs from our prior $p(D = d)$ in that values with higher relative likelihoods (such as PAIVS, for example) also have higher relative posterior probabilities. If someone has PAIVS, the likelihood they have left ventricular hypertrophy is 0.90, This extremely high likelihood is enough to make PAIVS the most likely of all six diseases given that a patient has LVH, even though PAIVS was only the fifth most likely disorder in our prior distribution $p(D = d)$.

(b.)

In order to compute the posterior $p(D = d|\text{LVH-report} = \text{yes})$, we need the prior $p(d)$ (already given to us), the likelihood $p(\text{LVH-report} = \text{yes}|D = d)$, and the marginal probability $p(\text{LVH-report=yes})$.

To compute the likelihood $p(\text{LVH-report} = \text{yes}|D = d)$, we need to sum over the two intermediate of LVH that can result in the measurement being true (corresponding to true positives and false positives). In other words, we have to compute

$$p(\text{LVH-report=yes|D=d}) =$$
$$p(\text{LVH-report=yes|LVH=yes})p(\text{LVH=yes|D=d}) +$$
$$p(\text{LVH-report=yes|LVH=no})p(\text{LVH=no|D=d}) =$$
$$0.90 * p(\text{LVH=yes|D=d}) + 0.05 * (1 - p(\text{LVH=yes|D=d}))$$

for each disease. These likelihoods are tabulated below.

We can simply sum up the **Prior×Likelihood** column to get the marginal probability.

Disease	Prior	Likelihood	Prior×Likelihood	Posterior
d	$p(D = d)$	$p(\text{Report}|D = d)$	$p(D = d)p(\text{Report}|D = d)$	$p(D = d|\text{Report})$
PFC	0.45	0.1350	0.06075	**0.34291**
TGA	0.14	0.1775	0.02485	**0.14026**
Fallot	0.03	0.152	0.00456	**0.02574**
PAIVS	0.06	0.815	0.04890	**0.27602**
TAPVD	0.12	0.0925	0.01110	**0.0627**
Lung	0.20	0.1350	0.02700	**0.15240**
Marginal		$p(\text{Report=yes})$	**0.17716**	**1**

The posterior probabilities are mostly the same as they were before we added in the test layer, but, most notably, PAIVS has slipped back into the second-most-likely disease given a positive test for LVH. This is because the uncertainty of the test layer reduces the likelihood of PAIVS. Its likelihood catches the full negative impact of the 10% chance of false positives (decreases from .90 to .81 and gets next to no increase in likelihood from the 5% chance of true negatives (.815).

Exercise 2.

a.

Suppose y has a Poisson distribution with mean θ and θ is gamma with shape parameter α and rate parameter β.

So $p(y|\theta) = \frac{1}{y!}\theta^y e^{-\theta}$
and
$p(\theta|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$.

By definition, the marginal distribution $p(y|\alpha,\beta)$ is the integral:

$$\int_0^\infty \delta\theta p(y|\theta)p(\theta|\alpha,\beta) =$$

$$\int_0^\infty \delta\theta \frac{1}{y!}\theta^y e^{-\theta}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} =$$

$$\int_0^\infty \delta\theta \frac{\beta^\alpha}{\Gamma(\alpha)y!}\theta^{y+\alpha-1}e^{-(\beta+1)\theta} \tag{3}$$

(Note that we only integrate over the positive reals for θ because the gamma function is not defined for the negative reals. On a related note, the Poisson distribution takes on discrete nonnegative values.)

After simplifying, we're faced with a difficult indefinite integral. However, given that functions in the form of distributions must integrate to one (and this looks particularly close to a gamma distribution), we might be able to find the marginal distribution simply by restructuring our integral into a known density times a function of α, β, and y.

While Equation 3 above appears to be very similar to the original gamma density, we still have to alter some of the parameters to make it work. For example, y! in the denominator must be rewritten as $\Gamma(y+1)$ or $y\Gamma(y)$. Now, given that the exponent on θ is now $y+\alpha-1$ and we have a denominator of $\Gamma(\alpha) * \Gamma(y)y$, we should probably be thinking about using a shape parameter of $(y+\alpha)$.

We also have $\Gamma(\alpha)\Gamma(y)y = B(\alpha, y)\Gamma(\alpha+y)y$, where $B(\alpha,\beta)$ denotes the beta function.

While this is getting quite complicated, recall that our integral is a function of *theta*, so that we just need to be able to put something inside the integral that will integrate to one.

In fact, we can use a similar trick to rewrite the exponent β^α as $\beta^{y+\alpha}\beta{-y}$ and remove the β^{-y} from the integral. Putting this all together:

$$\frac{1}{yB(\alpha,y)}\beta^{-y}\int_0^\infty \delta\theta \frac{\beta^{y+\alpha}}{\Gamma(\alpha+y)}\theta^{y+\alpha-1}e^{-(\beta+1)\theta} \tag{4}$$

Unfortunately there doesn't appear to be a good way to handle the exponential term at the end of Equation 4 unless we decide to change our rate parameter to $\beta+1$. The exponentiation e^θ can't be removed without some kind of complex transformation. So instead, let's just use a convoluted but effective multiplicative term of $\frac{(\beta+1)^{y+\alpha}}{\beta^{y+\alpha}}$ in the inside of the integral, while multiplying our outside terms by $\frac{\beta^{y+\alpha}}{(\beta+1)^{y+\alpha}}$.

This leaves us with

$$p(y|\alpha,\beta) = \frac{1}{yB(\alpha,y)}\beta^{-y}\frac{\beta^{y+\alpha}}{(\beta+1)^{y+\alpha}}\int_0^\infty \delta\theta \frac{(\beta+1)^{y+\alpha}}{\Gamma(y+\alpha)}\theta^{y+\alpha-1}e^{-(\beta+1)\theta}$$

$$= \frac{1}{yB(\alpha,y)}\frac{\beta^{\alpha}}{(\beta+1)^{y+\alpha}}\int_0^\infty \delta\theta\Gamma(y+\alpha,\beta+1) = \frac{\beta^{\alpha}}{(\beta+1)^{y+\alpha}B(\alpha,y)y}. \tag{5}$$

b.

Now suppose y_i (i = 1,2), y_1, y_2 independent has a Poisson distribution with mean $c\theta_i$ with $p(c) = \Gamma(\alpha,\beta)$.

What is the marginal distribution $p(y_1, y_2|\alpha,\beta,\theta_1,\theta_2)$? We have:
$p(y_i|c,\theta_i) = \frac{1}{y!_i}c^{y_i}\theta_i^{y_i}e^{-c\theta_i}$
$p(c|\alpha,\beta,\theta_i) = \frac{\beta^\alpha}{\Gamma(\alpha)}c^{\alpha-1}e^{-\beta c}$

$$p(y_i|\alpha,\beta,\theta_i) = \int_0^\infty \delta c\frac{1}{y_i!}c^{y_i}\theta_i^{y_i}e^{-c\theta_i}\frac{\beta^\alpha}{\Gamma(\alpha)}c^{\alpha-1}e^{-\beta c} \tag{6}$$

$$= \theta_i^{y_i}\int_0^\infty \delta c\frac{\beta^\alpha}{\Gamma(\alpha)y_i!}c^{y_i+\alpha-1}e^{-(\beta+\theta_i)c}$$

Once again, we seem to be being pushed towards a certain gamma distribution–in this case, $\Gamma(y_i+\alpha,\beta+\theta_i)$.

$$\Gamma(y_i+\alpha,\beta+\theta_i) = \frac{(\beta+\theta_i)^{y_i+\alpha}}{\Gamma(y_i+\alpha)}c^{y_i+\alpha-1}e^{-(\beta+\theta_i)c} \tag{7}$$

Notice that the last two multiplicative terms are already what we want in the final distribution.

Also note: As in part (a), we have $\Gamma(\alpha)\Gamma(y_i)y_i = B(\alpha,y_i)\Gamma(\alpha+y_i)y_i$, where $B(\alpha,\beta)$ denotes the beta function.

$$p(y_i|\alpha,\beta,\theta_i) = \theta_i^{y_i}\int_0^\infty \delta c\frac{\beta^\alpha}{\Gamma(\alpha)\Gamma(y_i)y_i}c^{y_i+\alpha-1}e^{-(\beta+\theta_i)c}$$

$$= \theta_i^{y_i}\int_0^\infty \delta c\frac{\beta^\alpha}{B(\alpha,y_i)\Gamma(y_i+\alpha)y_i}c^{y_i+\alpha-1}e^{-(\beta+\theta_i)c}$$

$$= \frac{{\theta_i}^{y_i}\beta^\alpha}{B(\alpha,y_i)(\beta+\theta_i)^{y_i+\alpha}y_i}\int_0^\infty \delta c\frac{(\beta+\theta_i)^{y_i+\alpha}}{\Gamma(y_i+\alpha)}c^{y_i+\alpha-1}e^{-(\beta+\theta_i)c}$$

$$= \frac{{\theta_i}^{y_i}\beta^\alpha}{B(\alpha,y_i)(\beta+\theta_i)^{y_i+\alpha}y_i}$$

$$\prod_{i=1,2} p(y_i|\alpha,\beta,\theta_i) = \prod_{i=1,2} \frac{\theta_i{}^{y_i}\beta^{\alpha}}{B(\alpha,y_i)(\beta+\theta_i)^{y_i+\alpha}y_i} \tag{8}$$

Exercise 3.

We know that Elvis had a twin brother. We want to find the probability, conditioned on knowing that Elvis had a twin brother, that this brother is an identical twin.

Suppose "tb" is the event that Elvis has a twin brother, "it" is the event that Elvis has an identical twin, and "ft" is the event that Elvis has a fraternal twin.

We have $p(tb|it) = 1$, because identical twins have the same sex, and $p(tb|ft) = .5$, because by assumption there's a 50% chance that a fraternal twin is of the same sex. We also have $p(it) = 1/300$ and $p(ft) = 1/125$, and we can find the marginal probability $p(tb) = p(tb|ft)p(ft) + p(tb|it)p(it) = .5 * 1/125 + 1/300 = 11/1500$ (assuming twins can only be fraternal or identical, and assuming that we can ignore correlation between twins, etc.).

We want to find $p(it|tb)$.

Applying Bayesian inference, where $\frac{\textbf{prior}\times\textbf{likelihood}}{\textbf{marginal}} = \textbf{posterior}$, we can calculate $p(it|tb) = \frac{p(it)\times p(tb|it)}{p(tb)} = \frac{1/300\times 1}{11/1500} = 5/11$.

Exercise 4.

Using the same reasoning and format as with Dr. X on the handout, we can simply compute the posterior in the standard way, and then we can compute the expected value of the posterior distribution to give the predictive probability that the next patient will be better, given each doctor's prior and the observed 18/21 success rate.

Below, Tables 2 and 3 demonstrate this calculation for both Dr. Y (0.8336) and Dr. Z (0.7908).

We can also use the tables to compute the posterior odds and the prior odds for each doctor.

Doctor	**Prior Odds**	**Posterior Odds**	**Bayes Factor**
	$g = P(H)/P(K)$	$f = P(H\|\text{data})/P(K\|\text{data})$	f/g
X	$\frac{6/11}{5/11} = 1.2$	$\frac{0.001}{0.999} \approx 0.001$	≈ 0.0008
Y	$\frac{15/55}{40/55} = .375$	$\frac{0.001}{0.999} \approx 0.001$	≈ 0.00229
Z	$\frac{0.95}{0.05} = 19$	$\frac{0.111}{0.888} \approx 0.125$	≈ 0.00657

Table 1: Computing the Bayes Factor for each doctor

Parameter	**Prior**	**Likelihood**	**Prior×Likelihood**	**Posterior**	$\theta\times$Posterior
θ	$p(\Theta = \theta)$	$p(\tilde{y} = B\|\Theta = \theta)$	$\Theta = \theta)p(\tilde{y} = B\|\Theta = \theta)$	$p(\Theta = \theta\|\tilde{y} = B)$	
0.0	0	0	0	0	0
0.1	$\frac{1}{55}$	tiny	tiny	tiny	tiny
0.2	$\frac{2}{55}$	tiny	tiny	tiny	tiny
0.3	$\frac{3}{55}$	tiny	tiny	tiny	tiny
0.4	$\frac{4}{55}$	tiny	tiny	tiny	tiny
0.5	$\frac{5}{55}$	4.77E-7	4.335E-8	tiny	tiny
0.6	$\frac{6}{55}$	6.50E-6	7.091E-7	0.0137	0.0082
0.7	$\frac{7}{55}$	4.40E-5	5.596E-6	0.108	0.0755
0.8	$\frac{8}{55}$	1.44E-4	2.096E-5	0.404	0.3233
0.9	$\frac{9}{55}$	1.50E-4	2.456E-5	0.473	0.4261
1.0	$\frac{10}{55}$	0	0	0	0
Marginal		$p(\tilde{y} = B)$	**5.1872E-5**	$p(y_{22} = B)$	**0.833568**

Table 2: Dr. Y

Parameter	**Prior**	**Likelihood**	**Prior×Likelihood**	**Posterior**	$\theta\times$Posterior
θ	$p(\Theta = \theta)$	$p(\tilde{y} = B\|\Theta = \theta)$	$\Theta = \theta)p(\tilde{y} = B\|\Theta = \theta)$	$p(\Theta = \theta\|\tilde{y} = B)$	
0.0	0.01	0	0	0	0
0.1	0.01	tiny	tiny	tiny	tiny
0.2	0.01	tiny	tiny	tiny	tiny
0.3	0.01	tiny	tiny	tiny	tiny
0.4	0.01	tiny	tiny	tiny	tiny
0.5	0.90	4.77E-7	4.292E-7	0.111	0.0554
0.6	0.01	6.50E-6	6.500E-8	0.0168	0.0101
0.7	0.01	4.40E-5	4.397E-7	0.113	0.0794
0.8	0.01	1.44E-4	1.441E-6	0.372	0.2974
0.9	0.01	1.50E-4	1.501E-6	0.387	0.34851
1.0	0.01	0	0	0	0
Marginal		$p(\tilde{y} = B)$	**3.876E-6**	$p(y_{22} = B)$	**0.790795**

Table 3: Dr. Z

Exercise 5.

First of all, let's use simple fraction algebra to alter the forms given for μ_1 and $\frac{1}{\tau_1^2}$:

$$\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2} = \frac{\sigma^2 + \tau_0^2}{\tau_0^2\sigma^2}$$

$$\mu_1 = \frac{\frac{\mu_0\sigma^2 + \tau_0^2 y}{\tau_0^2\sigma^2}}{\frac{\sigma^2+\tau_0^2}{\tau_0^2\sigma^2}} = \frac{\mu_0\sigma^2 + y\tau_0^2}{\sigma^2 + \tau_0^2}$$

Next, let's develop some of the forms that will appear in our derivation:

$$\mu_1^2 = (\frac{\mu_0\sigma^2 + y\tau_0^2}{\sigma^2 + \tau_0^2})^2$$

$$\frac{\mu_1^2}{\tau_1^2} = \frac{(\mu_0\sigma^2 + y\tau_0^2)^2}{(\sigma^2 + \tau_0^2)^2}\frac{\sigma^2 + \tau_0^2}{\tau_0^2\sigma^2} = \frac{(\mu_0\sigma^2 + y\tau_0^2)^2}{(\sigma^2 + \tau_0^2)\tau_0^2\sigma^2}$$

$$\frac{\mu_1}{\tau_1^2} = \frac{\mu_0\sigma^2 + y\tau_0^2}{\sigma^2 + \tau_0^2}\frac{\sigma^2 + \tau_0^2}{\tau_0^2\sigma^2} = \frac{\mu_0\sigma^2 + y\tau_0^2}{\tau_0^2\sigma^2} = \frac{\mu_0}{\tau_0^2} + \frac{y}{\sigma^2}$$

Now, let's calculate the likelihood times the prior.
We have:
$p(\theta) \propto \exp\frac{-(\theta-\mu_0)^2}{2\tau_0^2}$ and $p(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\frac{-(y-\theta)^2}{2\sigma^2}$

$$p(\theta|y) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp\frac{-(y-\theta)^2}{2\sigma^2}\exp\frac{-(\theta-\mu_0)^2}{2\tau_0^2}$$

$$p(\theta|y) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(y-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu_0)^2}{2\tau_0^2}\right)$$

$$p(\theta|y) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2\tau_0^2}\Big((\tau_0^2(\theta-y)^2 + \sigma^2(\theta-\mu_0)^2)\Big)\right]$$

$$p(\theta|y) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2\tau_0^2}\Big((\tau_0^2+\sigma^2)\theta^2 + \big(\mu_0^2\sigma^2 + \tau_0^2y^2\big) - 2\theta\big(\mu_0\sigma^2 + y\tau_0^2\big)\Big)\right]$$

$$p(\theta|y) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\theta^2}{2\tau_1^2} - \frac{1}{2\sigma^2\tau_0^2}\Big(\big(\mu_0^2\sigma^2 + \tau_0^2y^2\big) - 2\theta\big(\mu_0\sigma^2 + y\tau_0^2\big)\Big)\right]$$

$$p(\theta|y) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2}\left[\frac{\theta^2}{\tau_1^2} - \frac{2\theta\mu_1}{\tau_1^2} + \frac{\mu_0^2\sigma^2 + \tau_0^2y^2}{\sigma^2\tau_0^2}\right])$$

Because this is a posterior density whose random variable is θ, we can treat all other variables as constants, and in particular the third term in the exponential can for now be treated simply as a multiplicative constant of integration.

$p(\theta|y) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp{(-\frac{1}{2\tau_1^2}\Bigg[\theta^2 - 2\theta\mu_1 + \frac{\mu_0^2\sigma^2+\tau_0^2y^2}{\sigma^2+\tau_0^2}\Bigg])} \propto$
$\frac{1}{\sqrt{2\pi}\sigma} \exp{(-\frac{1}{2\tau_1^2}\Bigg[\theta^2 - 2\theta\mu_1\Bigg])} \propto$
$\frac{1}{\sqrt{2\pi}\sigma} \exp{(-\frac{1}{2\tau_1^2}\Bigg[\theta^2 - 2\theta\mu_1 + \mu_1^2\Bigg])} = \frac{1}{\sqrt{2\pi}\sigma} \exp{(-\frac{1}{2\tau_1^2}(\theta - \mu_1)^2)}$
$= \frac{1}{\sqrt{2\pi}\tau_1} \exp{(-\frac{1}{2\tau_1^2}(\theta - \mu_1)^2)}$

Because this posterior is a normal distribution (with posterior mean μ_1 and variance τ_1^2), which integrates to one (after we multiply by $\frac{\sigma}{\tau_1}$ in the last step of the above derivation), the marginal distribution $p(y) = \int_{-\infty}^{\infty} p(y|\theta)p(\theta)d\theta$ is just the inverse of the multiplicative constant we introduced above, now treated as a function of y:

$$p(y) = \frac{\tau_1}{\sigma} \exp\Bigg(- \frac{1}{2\tau_1^2}(\frac{\mu_0^2\sigma^2 + \tau_0^2y^2}{\sigma^2 + \tau_0^2} - \mu_1^2)\Bigg) \tag{9}$$

Of course, by the same argument as above, we can eliminate multiplicative terms that don't depend on y as constants of integration:

$$p(y) \propto \exp\Bigg(- \frac{1}{2\tau_1^2}(\frac{\tau_0^2y^2}{\sigma^2 + \tau_0^2})\Bigg) = \exp\Bigg(- \frac{y^2}{2\sigma^2}\Bigg) \tag{10}$$

The marginal distribution appears to be a normal centered around 0 with variance σ^2.

b.

The induction proof to a sample $\mathbf{y} = (y_1, ..., y_n)$ is mostly obvious once the base case has been established.

Just as we showed the conjugate transformations of (μ_0, τ_0) to (μ_1, τ_1) through the weighted averages below, we can do the same thing inductively to go from (μ_{n-1}, τ_{n-1}) to (μ_n, τ_n).

$$\mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \tag{11}$$

$$\tau_1 = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2} \tag{12}$$

The only thing to be careful about is deriving the sample mean $\bar{y}$ in the numerator of μ_n:

$$\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}\sum_{i=1}^{n-1} y_i + \frac{1}{\sigma^2}y_n = \frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}\sum_{i=1}^{n} y_i = \frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y} \tag{13}$$