

# Homework 2

Tim Vigers

September 16, 2020

## 1 BD 1.1.1

### 1. Example (a)

- (a) Here let  $X$  be a R.V. indicating the diameter of a pebble and  $Y = \log(X)$ . The logarithm of the diameter is normally distributed, so:

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

To find the distribution of  $X$ , we can do a simple transformation using  $\frac{d}{dx}Y = \frac{1}{X}$  and see that

$$P_X(X) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{\log(x)-\mu}{\sigma})^2}$$

- (b) Pebble diameters must be  $X > 0$ , so  $\log(X) \in \mathbb{R}$ . Because we are assuming  $\log(X) \sim \mathcal{N}(\mu, \sigma^2)$ ,  $\mu \in \mathbb{R}$  and  $\sigma > 0$ .
- (c) This is a parametric model because we are assuming a specific distribution for the pebble diameters.

### 2. Example (b)

- (a) For this example we have the model  $X_i = \mu + \epsilon_i$ , for  $1 \leq i \leq n$  and  $\epsilon \sim \mathcal{N}(0.1, \sigma^2)$ . Therefore

$$X_i \sim \mathcal{N}(\mu + 0.1, \sigma^2)$$

and

$$P_X(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu+0.1}{\sigma})^2}$$

- (b) In this case the variance of the errors is known, so the parameter space is  $\mu \in \mathbb{R}$ .

- (c) This is also a parametric model because we are assuming a distribution for the errors.

3. Example (c)

- (a) This is similar to the model above, but this time  $X_i = \mu + \epsilon_i$ , for  $1 \leq i \leq n$  and  $\epsilon \sim \mathcal{N}(\theta, \sigma^2)$ . Therefore

$$X_i \sim \mathcal{N}(\mu + \theta, \sigma^2)$$

and

$$P_X(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu+\theta}{\sigma}\right)^2}$$

- (b) The variance of the errors is still known, but this time we are only able to estimate the parameter  $\mu + \theta \in \mathbb{R}$  as the model is unidentifiable for  $\mu$  or  $\theta$  alone.
- (c) This is still a parametric model because we assume a distribution of the errors.

4. Example (d)

- (a) Let  $X$  = the number of eggs laid by an insect, which follows a Poisson distribution:

$$P_X(X) = \frac{e^{-\lambda} \lambda^x}{x!}$$

for  $x = 0, 1, \dots$  and  $\lambda > 0$ . If  $Y$  = the number of eggs that hatch assuming each egg hatches with probability  $p$ , then  $Y$  follows a binomial distribution given the number of eggs laid:

$$P_Y(Y|n = x) = \binom{x}{y} p^y (1 - p)^{x-y}$$

- (b)

$$\lambda > 0$$

$$Y = 0, 1, \dots$$

$$0 \leq p \leq 1$$

- (c) This is also a parametric model because we are assuming distributions for  $X$  and  $Y|X$ .

## 1.1 BD 1.1.2

1. Problem 1.1.1(c): It is possible to estimate the parameter  $\mu + \theta$ , but it is not possible to estimate  $\mu$  or  $\theta$  separately because there are many possible values of  $\mu$  and  $\theta$  that would produce the same  $\mu + \theta$ . For example,  $(\mu = 2, \theta = 2)$  and  $(\mu = 3, \theta = 1)$ .
2. The parameterization of 1.1.1(d) is indentifiable because the entomologist is collecting the number of eggs laid by each insect, which allows for estimation of  $\lambda$ . They are also collecting the number of eggs hatching, which makes it possible to estimate  $p$ .
3. Unlike the case above, if the entomologist is only collecting data on the number of eggs hatched, the model would be unidentifiable. The current parameterization assumes that  $n$  is known, so that if the entomologist records for example 6 eggs hatching out of a total of 36 eggs laid, they can estimate  $\hat{p} = \frac{1}{6}$ . However, if the number of eggs is unknown, then 6 hatchings could imply that  $\hat{p}_1 = \frac{6}{10}$ ,  $\hat{p}_1 = \frac{6}{6}$ , etc. because the denominator is unknown. Therefore,  $P_{\theta_1} = P_{\theta_2}$  does not imply  $\theta_1 = \theta_2$ .

## 1.2 BD 1.2.7

Example 1.1.1: Let  $X$  represent the number of defective items in a random sampling inspection where  $X(k) = k$  for  $k = 0, 1, \dots, n$ . If  $\theta$  represents the number of defective items in the population, then

$$p(X = k) = \frac{\binom{\theta}{k} \binom{N-\theta}{n-k}}{\binom{N}{n}}$$

Assume  $\theta$  has a  $\mathcal{B}(N, \pi_0)$  distribution:

$$\pi(\theta) = \binom{N}{\theta} \pi_0^\theta (1 - \pi_0)^{N-\theta}$$

Then we have that the posterior distribution of  $\theta$  given  $X = k$ :

$$\pi(\theta|X = k) = \frac{\pi(\theta)p(x|\theta)}{c} \propto \binom{N}{\theta} \pi_0^\theta (1 - \pi_0)^{N-\theta} \frac{\binom{\theta}{k} \binom{N-\theta}{n-k}}{\binom{N}{n}}$$

Where

$$c = \sum_{t=0}^n \pi(t)p(x|t) =$$

Blah blah blah, figure this part out...

$$\binom{N}{\theta} \pi_0^\theta (1 - \pi_0)^{N-\theta} \frac{\binom{\theta}{k} \binom{N-\theta}{n-k}}{\binom{N}{n}}$$

Which equals:

$$\frac{N!}{\theta!(N-\theta)!} \frac{(N-\theta)!}{(n-k)!(N-\theta-(n-k))!} \frac{N!}{\theta!(N-\theta)!} \frac{n!(N-n)!}{N!} \frac{\theta!}{k!(\theta-k)!} \pi_0^\theta (1 - \pi_0)^{N-\theta}$$

Several terms cancel, leaving us with:

$$\frac{n!(N-n)!}{k!(n-k)!(\theta-k)!(N-n-(\theta-k))!} \pi_0^\theta (1 - \pi_0)^{N-\theta}$$

This can be written as:

$$\binom{n}{k} \binom{N-n}{\theta-k} \pi_0^\theta (1 - \pi_0)^{N-\theta}$$

Multiplying this by  $\frac{\pi_0^k (1-\pi_0)^{n-k}}{\pi_0^k (1-\pi_0)^{n-k}}$  yields a constant  $\binom{n}{k} \pi_0^k (1 - \pi_0)^{n-k}$  and the kernel of a  $\mathcal{B}(N-n, \theta-k)$ :

$$\binom{N-n}{\theta-k} \pi_0^{\theta-k} (1 - \pi_0)^{N-n-(\theta-k)}$$

### 1.3 BD 1.2.12

1. Given  $X_1, \dots, X_n$  iid  $\mathcal{N}(\mu_0, \frac{1}{\theta})$  variables, the joint density  $p(x|\theta)$  is:

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \theta^{\frac{1}{2}} e^{-\frac{1}{2}\theta(x_i - \mu_0)^2} = \sqrt{2\pi}^{-n} \theta^{\frac{1}{2}n} e^{-\frac{n\theta}{2} \sum_{i=1}^n (x_i - \mu_0)^2}$$

Letting  $t = \sum_{i=1}^n (x_i - \mu_0)^2$ , this density is proportional to:

$$\theta^{\frac{1}{2}n} e^{-\frac{1}{2}\theta t}$$

2. If  $\pi(\theta) \propto \theta^{\frac{1}{2}(\lambda-2)} e^{-\frac{1}{2}\nu\theta}$ , then the posterior distribution

$$\pi(\theta|x) \propto \theta^{\frac{1}{2}(\lambda-2)} e^{-\frac{1}{2}\nu\theta} \theta^{\frac{1}{2}n} e^{-\frac{1}{2}\theta t}$$

by 1.2.10. This can be simplified to

$$\theta^{\frac{1}{2}(n+\lambda-2)} e^{-\frac{1}{2}\theta(\nu+t)} = \theta^{\frac{n+\lambda}{2}-1} e^{-\frac{\theta(\nu+t)}{2}}$$

which is the kernel of a

$$\text{Gamma}\left(\frac{n+\lambda}{2}, \frac{2}{\nu+t}\right) = \frac{1}{\Gamma\left(\frac{n+\lambda}{2}\right)\left(\frac{2}{\nu+t}\right)^{\frac{n+\lambda}{2}}} \theta^{\frac{n+\lambda}{2}-1} e^{-\frac{\theta(\nu+t)}{2}}$$

Using a simple change of variables where  $a = \theta(\nu+t)$ , this becomes:

$$\frac{1}{\Gamma\left(\frac{n+\lambda}{2}\right)2^{\frac{n+\lambda}{2}}} a^{\frac{n+\lambda}{2}-1} e^{-\frac{a}{2}}$$

So  $a \sim \chi_{n+\lambda}^2$ .

3. We can find the distribution of  $\sigma$  by plugging it into the posterior density:

$$p(\sigma|x) = \frac{1}{\Gamma\left(\frac{n+\lambda}{2}\right)\left(\frac{2}{\nu+t}\right)^{\frac{n+\lambda}{2}}} \left(\frac{2}{\sigma^3}\right) \left(\frac{1}{\sigma^2}\right)^{\frac{n+\lambda}{2}-1} e^{-\frac{\nu+t}{2\sigma^2}}$$

## 1.4 BD 1.3.8

1. To show that  $s^2$  is an unbiased estimator, we find its expected value:

$$E[s^2] = E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i^2 - n\bar{X}^2)\right] = \frac{1}{n-1} (nE[X_1^2] - nE[\bar{X}^2])$$

because the  $X_i$  are sampled from the same population.

$$\frac{1}{n-1} (nE[X_1^2] - nE[\bar{X}^2]) = \frac{1}{n-1} (n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)) = \frac{1}{n-1} (n\sigma^2 - \sigma^2)$$

This shows that  $E[s^2] = \sigma^2$  and it is therefore an unbiased estimator.

2. Because  $s^2$  is an unbiased estimator, the MSE is  $\text{Var}(s^2)$ . Using the fact that:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

it's obvious that

$$\text{Var}\left(\frac{(n-1)s^2}{\sigma^2}\right) = \text{Var}(\chi_{n-1}^2) = 2(n-1)$$

Rearranging this gives:

$$\text{Var}(s^2) = \frac{2\sigma^4(n-1)}{(n-1)^2} = \frac{2\sigma^4}{n-1}$$

3. If  $\hat{\sigma}_c^2 = c \sum_{i=1}^n (X_i - \bar{X})^2$ , then  $\hat{\sigma}_c^2 = c(n-1)s^2$ .

So,  $Var(\hat{\sigma}_c^2) = c^2(n-1)2\sigma^4$ . The bias of  $\hat{\sigma}_c^2$  is  $c(n-1)\sigma^2 - \sigma^2$ , so:

$$MSE(\hat{\sigma}_c^2) = c^2(n-1)2\sigma^4 + (c(n-1)\sigma^2 - \sigma^2)^2$$

Which expands to:

$$c^2(n-1)2\sigma^4 + c^2(n-1)^2\sigma^4 - 2c(n-1)\sigma^4 + \sigma^4$$

Taking the derivative with respect to  $c$  and setting equal to 0 gives:

$$4c(n-1)\sigma^4 + 2c(n-1)^2\sigma^4 - 2(n-1)\sigma^4 = 0$$

Dividing both sides by  $\sigma^4(n-1)$  results in:

$$4c + 2c(n-1) - 2 = 0$$

So

$$2c + c(n-1) = 1$$

and

$$c = \frac{1}{n+1}$$

To check that this is a minimum take the second derivative of the MSE:

$$\frac{d^2}{dc^2}MSE = 4(n-1)\sigma^4 + 2(n-1)^2\sigma^4 = 2\sigma^4(n-1)(n+1)$$

This function is positive for  $n > 1$ , so  $c = \frac{1}{n+1}$  minimizes MSE.