EXERCISES FOR SECTION 1 AND 2

Exercise 1.1 (Conditional probability). Suppose that if $\theta = 1$, then y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation σ. Also, suppose $Pr(\theta = 1) = 0.5$ and $Pr(\theta = 2) = 0.5$.

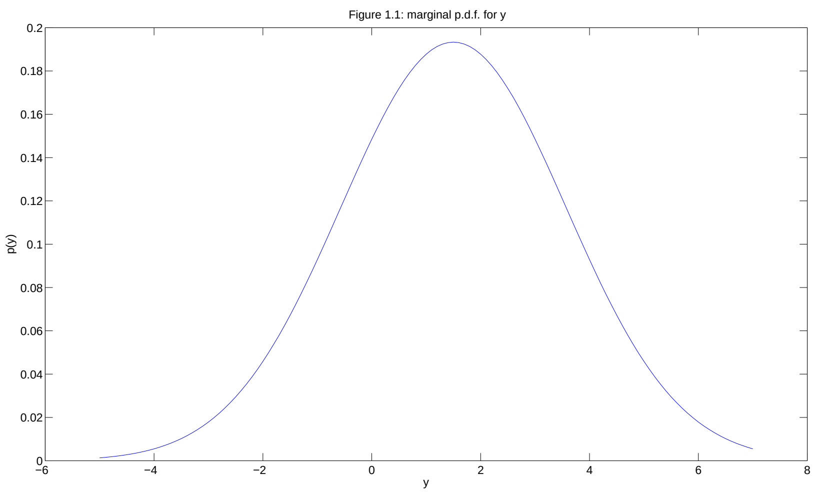
Two cases for θ:

$$\begin{aligned}
\theta = 1 &\Longrightarrow Pr(\theta = 1) = 0.5,\ \ y \sim N(1, \sigma^2) \\
\theta = 2 &\Longrightarrow Pr(\theta = 2) = 0.5,\ \ y \sim N(2, \sigma^2)
\end{aligned}$$

(a). For $\sigma = 2$, write the formula for the marginal probability density (marginal p.d.f.) for y and sketch it.

Solution. The marginal p.d.f. of y, $p(y)$, is given by

$$\begin{aligned}
p(y) = \sum_\theta p(y, \theta) &= \sum_\theta p(y|\theta)p(\theta) = p(y|\theta = 1)p(\theta = 1) + p(y|\theta = 2)p(\theta = 2) \\
&= N(y|1, 2^2)\frac{1}{2} + N(y|2, 2^2)\frac{1}{2} = \frac{1}{2}\left[N(y|1, 2^2) + N(y|2, 2^2)\right].
\end{aligned}$$



Bertil Wegmann, Department of Computer and Information Science, Linkping University SE-581 83 Linkping. E-mail: bertil.wegmann@liu.se.

(b). What is $Pr(\theta = 1|y = 1)$, again supposing $\sigma = 2$?

Solution.

$$p(\theta = 1|y = 1) = \frac{p(\theta = 1, y = 1)}{p(y = 1)} = \frac{p(y = 1|\theta = 1)p(\theta = 1)}{p(y = 1)}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\cdot 2}\exp\left[-\frac{1}{2\cdot 2^2}(1-1)^2\right]\cdot\frac{1}{2}}{\frac{1}{\sqrt{2\pi}\cdot 2}\cdot\frac{1}{2}\cdot\left[\exp\left[-\frac{1}{2\cdot 2^2}(1-1)^2\right]+\exp\left[-\frac{1}{2\cdot 2^2}(1-2)^2\right]\right]} = \frac{1}{1+\exp\left[-\frac{1}{8}\right]} \approx 0,53$$

.

(c). Describe how the posterior density of θ changes in shape as σ is increased and as it is decreased.

Solution. The posterior density is given by

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{\exp\left[-\frac{1}{2\sigma^2}(y-\theta)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}(y-1)^2\right]+\exp\left[-\frac{1}{2\sigma^2}(y-2)^2\right]}.$$

The posterior probability of $\theta = 1$ is given by

$$p(\theta = 1|y) = \frac{1}{1+\exp\left[\frac{1}{2\sigma^2}\left[(y-1)^2-(y-2)^2\right]\right]} = \frac{1}{1+\exp\left[\frac{2y-3}{2\sigma^2}\right]}.$$

Similarly,

$$p(\theta = 2|y) = \frac{1}{1+\exp\left[-\frac{2y-3}{2\sigma^2}\right]}.$$

Thus,

$$\sigma^2 \longrightarrow \infty \Longrightarrow p(\theta|y) \longrightarrow p(\theta) = \frac{1}{2}$$

$$\sigma^2 \longrightarrow 0 \Longrightarrow 2 \text{ scenarios}$$

$$y < \frac{3}{2}, \sigma^2 \longrightarrow 0 \Longrightarrow p(\theta = 1|y) \longrightarrow 1.$$

$$y > \frac{3}{2}, \sigma^2 \longrightarrow 0 \Longrightarrow p(\theta = 2|y) \longrightarrow 1.$$

Exercise 1.6 (Conditional probability). approximately 1/125 of all births are fraternal twins and 1/300 of births are identical twins. Elvis Presley had a twin brother (who died at birth). What is the probability that Elvis was an identical twin? (You may approximate the probability of a boy or girl birth as $\frac{1}{2}$.)

Solution. Events: A = Elvis had a twin brother, B = Elvis was an identical twin, C = Elvis was a fraternal twin. $P(boy) = P(girl) = \frac{1}{2}$ gives

$$P(B) = \frac{1}{2}\frac{1}{300} = \frac{1}{600}, \;\; P(C) = \frac{1}{2}\frac{1}{125} = \frac{1}{250}.$$

This gives,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|C)P(C)} = \frac{1\frac{1}{600}}{1\frac{1}{600} + \frac{1}{2}\frac{1}{250}} = \frac{5}{11}.$$

Exercise 2.1 (Posterior inference). suppose there is $Beta(4, 4)$ prior distribution on the probability θ that a coin will yield a 'head' when spun in a specified manner. The coin is independently spun ten times, and 'heads' appear fewer than 3 times. You are not told how many heads were seen, only that the number is less than 3. Calculate your exact posterior density (up to a proportionality constant) for θ and sketch it.

Solution. The prior distribution for θ is

$$p(\theta) \propto \theta^3(1-\theta)^3$$

Let y = total number of heads in n spuns. Then,

$$y|\theta \sim Bin(n = 10, \theta) \Longrightarrow p(y|\theta) = \binom{10}{y}\theta^y(1-\theta)^{10-y}.$$

We have that

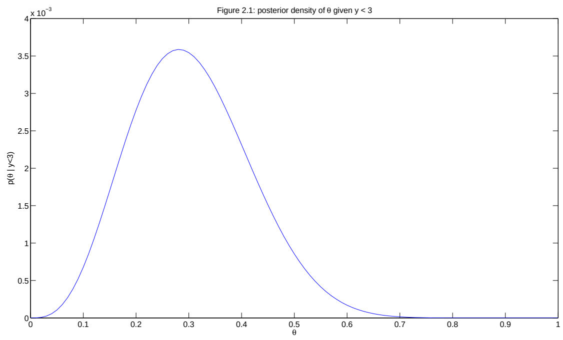
$$p(y < 3|\theta) = \sum_{y=0}^{2} p(y|\theta) = (1-\theta)^{10} + 10\theta(1-\theta)^9 + 45\theta^2(1-\theta)^8,$$

so that

$$p(\theta|y < 3) = \frac{p(\theta, y < 3)}{p(y < 3)} = \frac{p(y < 3|\theta)p(\theta)}{p(y < 3)} \propto p(y < 3|\theta)p(\theta).$$

This gives

$$p(\theta|y < 3) \propto \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11}.$$



Exercise 2.5 (posterior distribution as a compromise between prior information and data). Let y be the number of heads in n spins of a coin, whose probability of heads is θ.

(a). If your prior distribution for θ is uniform on the range $[0, 1]$, derive your prior predictive distribution for y,

$$Pr(y = k) = \int_0^1 Pr(y = k|\theta)d\theta,$$

for each $k = 0, 1, \dots, n$.

Solution. $y \sim Bin(n, \theta) \Longrightarrow p(y = k|\theta) = \binom{n}{k}\theta^y(1-\theta)^{n-y}$, so

$$p(y = k) = \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} d\theta.$$

If $\theta \sim \text{Beta}(k + 1, n - k + 1)$, then

$$1 = \int_0^1 p(\theta) = \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} \int_0^1 \theta^k(1-\theta)^{n-k} d\theta.$$

This gives that

$$p(y = k) = \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{n!}{k!(n-k)!} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}.$$

(b). Suppose you assign a $\text{Beta}(\alpha, \beta)$ prior distribution for θ, and then you observe y heads out of n spins. Show algebraically that your posterior mean of θ always lies between your prior mean, $\frac{\alpha}{\alpha+\beta}$, and the observed relative frequency of heads, $\frac{y}{n}$.

Solution.

$$\theta \sim \text{Beta}(\alpha, \beta) \Longrightarrow p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$
$$y|\theta \sim Bin(n, \theta) \Longrightarrow p(y|\theta) \propto \theta^y(1-\theta)^{n-y}.$$

This gives

$$p(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \propto \text{Beta}(\theta|\alpha + y, \beta + n - y).$$

Hence,

$$E[\theta|y] = \frac{\alpha + y}{\alpha + \beta + n} = \frac{\alpha}{\alpha+\beta} \cdot \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right) + \frac{y}{n} \cdot \left(\frac{n}{\alpha+\beta+n}\right).$$

(c). Show that, if the prior distribution on θ is uniform, the posterior variance of θ is always less than the prior variance.

Solution. We have that

$$p(\theta|y) \propto \theta^y(1-\theta)^{n-y} \propto \text{Beta}(\theta|y+1, n-y+1).$$

This gives

$$Var(\theta|y) = \frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} \leq \frac{(\frac{n}{2}+1)(n-\frac{n}{2}+1)}{(n+2)^2(n+3)} \leq \frac{1}{16} < \frac{1}{12} = Var(\theta).$$

(d). Give an example of a Beta(α, β) prior distribution and data y, n, in which the posterior variance of θ is higher than the prior variance.

Solution.

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \text{ and } Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$p(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \propto \text{Beta}(\theta|\alpha+y, \beta+n-y) \text{ and}$$
$$Var(\theta|y) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}.$$

For example, $n = 2, y = 1, \alpha = 1, \beta = 9$ gives $Var(\theta) = \frac{9}{1100} < \frac{10}{936} = Var(\theta|y)$.

Exercise 2.8 (normal distribution with unknown mean). a random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\bar{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean θ and known standard deviation 20 pounds. Suppose your prior distribution for θ is normal with mean 180 and standard deviation 40.

(a). Give your posterior distribution for θ. (Your answer will be a function of n.)

Solution.

$$\bar{y}|\theta \sim N\left(\theta, \frac{20^2}{n}\right)$$
$$\theta \sim N(180, 40^2).$$

Equation (2.11) – (2.12) gives

$$p(\theta|\bar{y}) = N(\theta|\mu_n, \tau_n^2),$$

where

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}},$$

and

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}.$$

(b). A new student is sampled at random from the same population and has a weight of $\tilde{y}$ pounds. Give a posterior predictive distribution for $\tilde{y}$.

Solution. The posterior predictive distribution is given by (page 47-48)

$$p(\tilde{y}|\bar{y}) = \int_{-\infty}^{\infty} p(\tilde{y}|\theta)p(\theta|\bar{y})d\theta \propto \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2\right] \exp\left[-\frac{1}{2\tau_n^2}(\theta-\mu_n)^2\right] d\theta.$$

This gives that $\tilde{y}|y$ is normal distributed with

$$E(\tilde{y}|\bar{y}) = E[E(\tilde{y}|\theta,\bar{y})|\bar{y}] = E[\theta|\bar{y}] = \mu_n$$

and

$$Var[\tilde{y}|\bar{y}] = E[Var(\tilde{y}|\theta,\bar{y})|\bar{y}] + Var[E(\tilde{y}|\theta,\bar{y})|\bar{y}] = E[\sigma^2|\bar{y}] + Var(\theta|\bar{y}) = \sigma^2 + \tau_n^2.$$

(c). For $n = 10$, give a 95% posterior interval for θ and a 95% predictive interval for $\tilde{y}$.

Solution. A 95% posterior interval for θ is given by

$$E[\theta|\bar{y} = 150] \pm 1,96\sqrt{Var(\theta|\bar{y} = 150)}.$$

This gives the interval to

$$150,73 \pm 1,96\sqrt{39,024}.$$

A 95% posterior predictive interval for $\tilde{y}$ is given by

$$E(\tilde{y}|\bar{y}) \pm 1,96\sqrt{Var[\tilde{y}|\bar{y}]}.$$

This gives the interval to

$$150,73 \pm 1,96\sqrt{439,024}.$$

(d). For $n = 100$, give a 95% posterior interval for θ and a 95% predictive interval for $\tilde{y}$.

Solution. The 95% posterior interval for θ becomes $[146, 154]$.

The 95% posterior predictive interval for $\tilde{y}$ becomes $[111, 189]$.

Exercise 2.11. Figure for part (a):