Example: Inference about a genetic status (Sec. 1.4, pages 8-9)

- Hemophilia is a disease that mainly affects males
- The reason is that males have one X-chromosome and one Y-chromosome, while females have two X-chromosomes
- The Y-chromosome is passed on from the father, while the X-chromosome (for males) was passed on from the mother
- The disease exhibits X-chromosome-linked recessive inheritance, which means that males are affected if their one X-chromosome carries the gene, while females are affected only if both their X-chromosomes carry the gene
- A female that carries the gene on one of her X-chromosomes, is said to be a carrier of the disease, because she can potentially pass on the gene to her children.

Prior distribution

- Consider a woman that has an affected brother and an unaffected father, and she does not have the disease
  $\Rightarrow$ Her mother was a carrier, and the woman has probability 0.5 of having inherited the gene and being a carrier

sep 1-12:20

- Let θ be the unknown indicator of whether she is a carrier ($\theta=1$) or not ($\theta=0$).
- Hence, based on the prior knowledge, the prior distribution for θ is $p(\theta=1)=p(\theta=0)=0.5$

Sampling distribution

- The data from which we are going to make posterior inference are whether her sons are affected or not
- Let $y_i=1$ if son $\underline{i}$ is affected, and $y_i=0$ if he is not affected
- We assume that none of her sons are identical twins, hence the outcomes of the sons can be assumed exchangeable and independent conditional on θ.

  Hence $p(y_1,\dots,y_n|\theta)=\prod_{i=1}^{n} p(y_i|\theta)$

  and $p(y_i|\theta)$ is given by:

  $Pr(y_i=0|\theta=1)=0.5$ $\quad$ $Pr(y_i=0|\theta=0)=1$

  $Pr(y_i=1|\theta=1)=0.5$ $\quad$ $Pr(y_i=1|\theta=0)=0$

sep 1-12:32

Data and likelihood

- Suppose she has 2 sons, and neither of them are affected. Hence, the likelihood is given by

$$Pr(y_1=0, y_2=0 \mid \theta=1) = 0.5 \cdot 0.5 = 0.25$$
$$Pr(y_1=0, y_2=0 \mid \theta=0) = 1 \cdot 1 = 1$$

Posterior distribution of θ

We now combine the prior distribution with the information from the data (through the likelihood) to get the posterior distribution of θ, i.e. the posterior probability that the woman is a carrier:

$$Pr(\theta=1 \mid y) = \frac{Pr(\theta=1) \cdot Pr(y_1=0, y_2=0 \mid \theta=1)}{\sum_{j=0,1} Pr(\theta=j) \cdot Pr(y_1=0, y_2=0 \mid \theta=j)} = \frac{0.5 \cdot 0.25}{0.5 \cdot 0.25 + 0.5 \cdot 1} = 0.2$$

$$Pr(\theta=0 \mid y) = 1 - Pr(\theta=1 \mid y) = 0.8$$

sep 1-12:41

Prediction

- $p(\theta|y)$ gives posterior conclusions regarding the unobservable parameter θ
- We may also want to make inference on the unobserved, but potentially observable, $\tilde{y}$ (for example future observations)

This we call predictive inference

We of course rely on the same assumptions

Before we have any data, the marginal distribution of the unknown, but observable, y is (θ continuous)

$$p(y) = \int p(y,\theta)\, d\theta = \int p(\theta)\, p(y|\theta)\, d\theta$$

In this context, we call it the prior predictive distribution

prior to observing any data — The distribution of a quantity that is observable

After we have observed the data y, we can make predictive inference for an unknown, observable $\tilde{y}$ from the same model, conditional on y.

sep 1-12:53

Predictive inference is then made in terms of the posterior predictive distribution:

posterior — Condition on previously observed data

predictive — Prediction of an observable $\tilde{y}$

$$p(\tilde{y} \mid y) = \int p(\tilde{y}, \theta \mid y)\, d\theta = \int p(\tilde{y} \mid \theta, y) \cdot p(\theta \mid y)\, d\theta = \int \underbrace{p(\tilde{y} \mid \theta)}_{\text{The sampling distribution of } \tilde{y} \text{ given } \theta} \cdot \underbrace{p(\theta \mid y)}_{\text{Posterior distribution of } \theta \text{ given } y}\, d\theta$$

θ discrete

$$p(\tilde{y} \mid y) = \sum_{\theta} p(\tilde{y} \mid \theta) \cdot p(\theta \mid y)$$

sep 1-13:00

Genetic status example continued:

- Suppose the woman is pregnant with her third son, whose hemophilia status is yet unknown. We wish to make predictive inference regarding whether this son is affected or not, given that her first two sons were not. Let $\tilde{y}_3$ be the unobserved status of the third son

Now $\Pr(\tilde{y}_3 = 0 \mid \theta = 1) = 0.5$ $\quad \Pr(\tilde{y}_3 = 0 \mid \theta = 0) = 1$

$\Pr(\tilde{y}_3 = 1 \mid \theta = 1) = 0.5$ $\quad \Pr(\tilde{y}_3 = 1 \mid \theta = 0) = 0$

$$\Pr(\tilde{y}_3 = 1 \mid y_1 = 0, y_2 = 0) = \sum_{j \in \{0,1\}} \Pr(\tilde{y}_3 = 1 \mid \theta = j) \cdot \Pr(\theta = j \mid y_1 = 0, y_2 = 0)$$

$$= 0 \cdot 0.8 + 0.5 \cdot 0.2 = \underline{\underline{0.1}}$$

sep 1-13:18

Sequential analyses (adding new data)

- A useful feature of Bayesian inference is that it is straightforward to update the posterior distribution with new data, which enables sequential analyses.
- We follow the same recepy, and simply treat the current posterior distribution as the new prior, and update by combining it with the likelihood of the new data to get the new posterior distribution
- Let y be the data already included in the current posterior distribution, and y^{new} the new, additional data. Then the new, updated posterior distribution is obtained by using Bayes' formula (again):

$$p(\theta | y, y^{new}) = \frac{p(\theta, y^{new} | y)}{p(y^{new}|y)} = \frac{p(\theta|y) \cdot p(y^{new}|\theta, y)}{p(y^{new}|y)}$$

$$\underset{\substack{y \text{ and } y^{new} \\ \text{are cond. indep. given } \theta}}{=} \frac{p(\theta|y)\cdot p(y^{new}|\theta)}{p(y^{new}|y)}$$

$$p(y^{new}|y) \rightarrow \begin{cases} \sum_\theta p(\theta|y) \cdot p(y^{new}|\theta), & \theta \text{ discrete} \\ \int p(\theta|y) \cdot p(y^{new}|\theta), & \theta \text{ continuous} \end{cases}$$

Genetic status example continued:

- Suppose the womans third son is unaffected. Now

$$Pr(\theta = 1 \mid y_1 = 0, y_2 = 0, y_3 = 0) = \frac{Pr(\theta=1|y_1=0, y_2=0) \cdot Pr(y_3=0|\theta=1)}{\sum_{j \in \{0,1\}} Pr(\theta=j|y_1=0,y_2=0) \cdot Pr(y_3=0|\theta=j)}$$

$$= \frac{0.2 \cdot 0.5}{0.2 \cdot 0.5 + 0.8 \cdot 1} = 0.111$$

With data $y_1=0, y_2=0$: $Pr(\theta=1 \mid y_1=0, y_2=0) = 0.2$

Added $y_3 = 0$: $Pr(\theta=1 \mid y_1=0, y_2=0, y_3=0) = 0.111$

Conversely, if $y_3 = 1$:

$Pr(y_3=1|\theta=0)=0$ and hence $Pr(\theta=1|y_1=0, y_2=0, y_3=1) = \frac{0.2 \cdot 0.5}{0.2 \cdot 0.5} = \underline{\underline{1}}$

Probability and uncertainty

- Classical inference: Speak of probability in relation to outcomes of experiments that in theory can be repeated (independent and identically) infinitely many times. The probabilities are then relative frequencies of events. Hence, the name frequentist inference
    - Simple example: The probability that a fair toss of a fair coin results in heads (observable and repeatable)

- Bayesian inference: Probability is interpreted in a broader sense, as a fundamental way of quantifying and communicating uncertainty
    - Hence, probability can be used to make statements about
        - observable outcomes of repeatable experiments (as for classical inference)
        - BUT also unobservable or unrepeatable phenomenae, parameters etc

sep 1-13:45

Examples: The probability of rain tomorrow
(observable, but not repeatable)

The probability of a tie in a football match tomorrow
(observable, but not repeatable)

The probability of a tier in tomorrows match given that it does not rain tomorrow
(not repeatable, unobservable if it does not rain tomorrow)

In Bayesian analysis, knowledge about anything unknown is a described by a probability distribution.
But this can of course be challenging to formulate well.

sep 1-13:54

Summarizing inferences by simulation

- Simulation/sampling from (posterior) distributions is an important tool in Bayesian inference
  - ↳ Often "easy" to draw samples when the density is not tractable analytically
- ↳ When the sample is large enough, the histogram of the random draws from the distribution gets close to the true density, also sample moments, percentiles etc provide good estimates of different aspects of the distribution, and the precision of these estimates can be estimated.

  Example: To estimate the 75th percentile of the posterior distribution of θ, draw a random sample of size S from $p(\theta|y)$ and compute the 95th percentile order statistic from the samples.
- ↳ Sampling from a probability distribution is straightforward within e.g. R

sep 1-13:59

Table 1.1

Simulation draw	Parameters				Predictive quantities				$\frac{\theta_1}{\theta_3}$	$\tilde{y}_1+\tilde{y}_2$
	θ_1	θ_2	$\cdots$	θ_k	$\tilde{y}_1$	$\tilde{y}_2$	$\cdots$	$\tilde{y}_n$		
1	θ_1^1	θ_2^1	$\cdots$	θ_k^1	$\tilde{y}_1^1$	$\tilde{y}_2^1$	$\cdots$	$\tilde{y}_n^1$	$\frac{\theta_1^1}{\theta_3^1}$	
2	$\vdots$	$\vdots$		$\vdots$	$\vdots$			$\vdots$	$\frac{\theta_1^2}{\theta_3^2}$	
$\vdots$									$\vdots$	
S	θ_1^s	θ_2^s	$\cdots$	θ_k^s	$\tilde{y}_1^s$	$\tilde{y}_2^s$	$\cdots$	$\tilde{y}_n^s$	$\frac{\theta_1^s}{\theta_k^s}$	

sep 1-14:25