Exercises for 'Introduction to Bayesian Statistics'

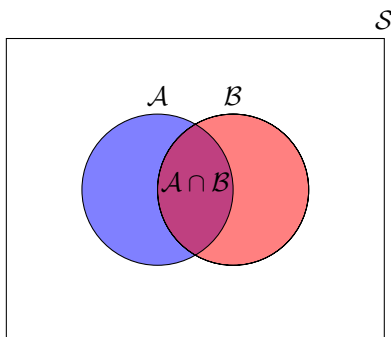## Probability

1. Axioms

$$P(\mathcal{S}) = 1,\ P(\emptyset) = 0$$

$$P(\bigcup_{i=1}^{\infty} \mathcal{A}_i) = \sum_{i=1}^{\infty} \mathcal{A}_i$$

From these two axioms, three important properties can be deduced.

$$P(\mathcal{A}^c) = 1 - P(\mathcal{A}^c)$$
$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$$
$$P(\mathcal{A}) \leq P(\mathcal{B}) \quad \text{if } \mathcal{A} \subset \mathcal{B}$$

Using the Venn diagram below, show that

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$$

2. Marginal and Conditional Probability

Consider the following two-dimensional matrix:

```
##        blond brown  red black
## blue    0.22  0.21 0.00  0.01
## green   0.00  0.14 0.06  0.01
## brown   0.16  0.15 0.00  0.04
```

    a. Calculate the marginal probabilities.
    b. Calculate the conditional probability of eye color given blond hair.

3. Intuitions about (conditional) probability

Imagine that there are three cards: one is red on either side, the other is white on either side, and the third is red on one side and white on the other. Suppose a confederate draws a card from this set of three totally at random, and shows a totally random side of that card to you.

a. What is the probability that what you see is red?
b. What is the probability that, given that you see red, the other side of the card that the confederate holds is white?
c. Write down a two-dimensional probability matrix to support your answers.
d. Now imagine that the confederate draws a card at random but then presents "red" with probability $\frac{1}{4}$ if there is a red side. What is the conditional probability that the other side of the card is white if you are shown red? (Best use a new two-dimensional probability matrix in support of your answer.)

Random variables

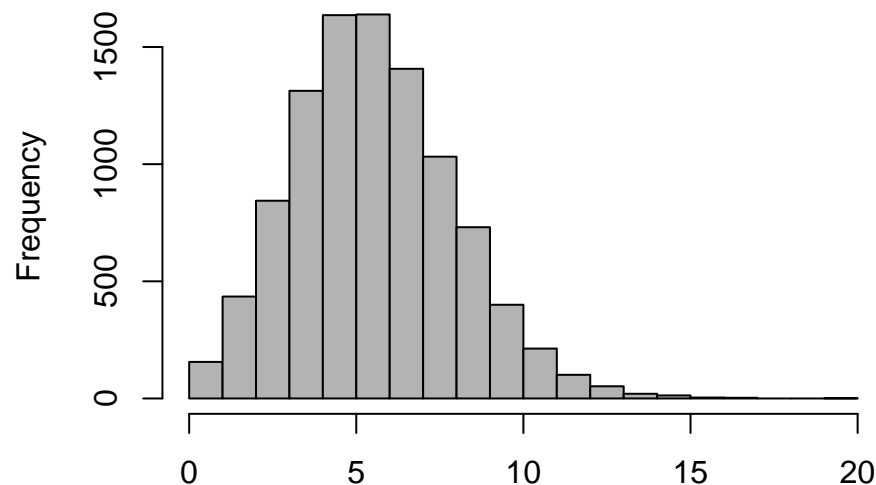1) By hand, compute the mode, median, mean, and the standard deviation of the following values

```
##  [1] 6 0 5 7 6 4 4 4 5 7
```

Distributions

1. Guessing game

For this exercise you should consult Wikipedia.

a. See the following distribution below. Do you think this variable is Poisson, Exponential, Gaussian, Binomial, or Gamma distributed?

b. For each of the options, list your reasons for why you chose, or not chose, it as an answer.
c. How do these distributions differ? How do they arrive in nature? Give an example for each distribution!

Classical and Bayesian statistics

1. Overview

Make a table listing the differences and commonalities of classical and Bayesian statistics.

2. Classical concepts

a. Define the *p*-value.

    b. Define a confidence interval, give a practical example.
    c. What is statistical power? How do you determine it?

3. Bayesian concepts

    a. What is Bayes' rule? How do you derive it?
    b. What is the difference between prior, likelihood, and posterior?
    c. What is a Bayes factor? How do you compute it?

4. Coin flips

Given three coin flips, all coming up heads, what is the probability that the next coin flip will show heads? Give the prediction via maximum likelihood and – assuming an uninformative prior – Bayesian inference (use the mean, median, and mode of the posterior). Discuss your findings.

5. A night at the club

Jack & Jill debate whether they should enter club Whatever tonight. They only want to go if the women/men ratio is about even. Otherwise the sphere just isn't, like, right, you know. Before they decide to pay, they screen the crowd that is lining up at the entrance. It is now 1:30 and there's 28 people in line, of which 9 female, 16 men, and three unidentifiable. How should they test whether the women/men ratio is even in this particular case? Should they use estimation, model comparison, or p-value testing? If you cannot decide, discuss pros and cons for each. (Think about how they should construe a likelihood function, whether a reasonable prior is at hand, whether the sampling procedure is reasonably clear, whether model complexity plays a role, etc. Suppose for the sake of argument, that computational complexity is not a problem: Jack & Jill have brought their laptops, of course.)

Statistical modeling

1. Coin flips

Is the coin fair, or is it not? Assume you observe the data

```
## [1] 1 1 1 1 1
```

    a. Draw the Binomial graphical model and your prior specification.
    b. What is the maximum likelihood estimate for θ, given the data? Is it a reasonable estimate? Why, why not?
    c. Compute the posterior distribution using Bayes' rule.
    d. What is the mean of this posterior distribution? Contrast this Bayesian estimate with the maximum likelihood estimate.
    e. Show that the mode of the posterior distribution under an uninformative prior is equivalent to the maximum likelihood estimate.

2. Univariate Gaussian, Variance known

Assume you want to test whether an IQ score of a prison inmate differs from 100. You assume you know the standard deviation; the only thing you want to infer is the mean, μ. You observe the following data

```
## [1] 101 112 107
```

a. Draw the Gaussian graphical model specifying the statistical model. What are your priors?
b. Sketch how you would arrive at the posterior distribution.

3. Multivariate Gaussian

Assume X is a bivariate Gaussian random variable with $\mu = (0,0)$ and $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Draw the graphical model using an uninformative prior over ρ.

Data Analysis with JASP

1. Correlation

a. Load *Presidents.csv* into JASP. Run descriptive statistics and summarize what you find.
b. Conduct a classical correlation test. Interpret the resulting p-value and confidence interval. Are these concepts valid for the current data set? *Hint:* think about the data generating process.
c. Conduct a Bayesian correlation test. Estimate the posterior distribution and interpret the Bayes factor. Does the default prior in JASP make sense?

1. t-test

a. Load *KitchenRolls.csv* into JASP. Run descriptive statistics and summarize the data graphically.
b. Conduct a classical t-test. Interpret the resulting p-value and confidence interval. Are these concepts valid for the current data set? *Hint:* think about the data generating process.
c. Conduct a Bayesian t-test. Estimate the posterior distribution and interpret the Bayes factor. Does the default prior in JASP make sense?
d. Conduct a Bayesian one-sided t-test. Is this approach justifiable?
e. Run a robustness check and a sequential analysis. What do you conclude?