MS2505: Bayesian Statistics
Compiled exercises

December 13, 2024

This document contains a compilation of all suggested exercises for the course MS2505: Bayesian statistics at Blekinge Institute of Technology[1].

1 Repetition

Basic probability theory notation and terms: This can be trivial or you may need to refresh your memory on these concepts. Note that some terms may be different names for the same concept. Explain each of the following terms with one sentence:

- probability
- probability mass
- probability density
- probability mass function (pmf)
- probability density function (pdf)
- probability distribution
- discrete probability distribution
- continuous probability distribution
- cumulative distribution function (cdf)
- likelihood

2 Repetition Exercises 1

1. Suppose a new genetic test can detect a rare mutation with 95% sensitivity and 99% specificity. The mutation occurs in 1/1000 of the population. A person tests positive for the mutation.

   $a)$ What is the prior probability that the person has the mutation?

   $b)$ What is the likelihood of observing a positive test result given the person has the mutation?

   $c)$ Compute the posterior probability that the person has the mutation given the positive test result.

2. A pharmaceutical company is testing a new drug that works for some genetic subtypes of a disease. From previous studies, it is believed that 70% of patients respond to the drug if they carry the genetic marker (θ = 1), and only 10% of patients respond if they do not have the marker (θ = 0). Suppose 20% of the population has the marker. A patient is given the drug and responds positively.

   $a)$ Write down the prior probabilities for θ = 1 and θ = 0.

   $b)$ Derive the likelihood of the observed response for each possible θ.

   $c)$ Compute the posterior probability that the patient has the genetic

3. An online retailer is trying to model the probability that a customer will make a purchase based on their browsing behavior. Let θ denote the unknown probability that a visitor makes a purchase. Suppose the prior distribution of θ is $Beta(\alpha = 2, \beta = 3)$, and the retailer observes that 3 out of 10 visitors made purchases.

a) Derive the posterior distribution of θ.

b) What is the expected value of θ under the posterior distribution?

c) If the retailer observes an additional 2 purchases from 5 new visitors, update the posterior distribution and compute the new expected value of θ.

4. Suppose that if θ = 0, then y follows a geometric distribution with success probability p = 0.4, and if θ = 1, y follows a negative binomial distribution with parameters r = 2 and p = 0.3. Assume that $\Pr(\theta = 0) = 0.6$ and $\Pr(\theta = 1) = 0.4$.

   a) Derive the formula for the marginal probability mass function of y.

   b) Given y = 3, calculate $\Pr(\theta = 1|y = 3)$.

5. Suppose y follows a binomial distribution for given n and unknown parameter θ, where the prior distribution of θ is $Beta(\alpha, \beta)$. Find the marginal distribution of y.

3 Repetition exercises 2

1. A diagnostic test has a sensitivity (true positive) of 98% and a specificity (true negative) of 95%. The prevalence of the disease is 1 in 500. A person tests positive.

   a) Compute the prior probability that the person has the disease.

   b) Find the probability of observing a positive test result if the person does not have the disease.

   c) Using Bayes' theorem, calculate the posterior probability that the person has the disease given the positive test result.

2. Suppose an online platform models the probability of users clicking on an ad. Let the prior distribution of the click-through rate θ be $Beta(2, 2)$. After observing 8 clicks out of 20 impressions:

   (a) Derive the posterior distribution of θ.

   (b) Compute the posterior mean of θ.

   (c) Predict the probability of observing 5 clicks out of the next 10 impressions.

3. Suppose θ has a prior $Beta(3, 5)$ and the observed data y is modeled as $Binomial(n = 10, \theta)$.

   a) Derive the marginal distribution of y.

   b) Compute the posterior distribution of θ if $y = 6$.

   c) Calculate the posterior predictive probability of $y = 8$ in a new trial.

4. Consider a normal distribution with known variance $\sigma^2 = 4$. Suppose the prior distribution for the mean μ is $N(10, 9)$. After observing data $y = [12, 14, 11]$:

   a) Derive the posterior distribution of μ.

   b) Compute the posterior predictive mean and variance of a new observation $\tilde{y}$.

   c) Find the probability that $\tilde{y} > 15$.

5. Mark is deciding whether to take the highway or a backroad on his trip to a nearby town. He knows that the highway is usually faster, but sometimes it can be congested due to construction or accidents. Mark observes that there is heavy traffic reported on the highway, and he needs to determine if the congestion is a one-time event or if it will continue for the day. Based on the prior information and the observed traffic delay, Mark applies Bayesian decision theory to decide whether to take the highway or the backroad.

   a) Mark knows the following prior information:

      - The prior probability of construction causing a delay on the highway is 10% (i.e., $P(Construction) = 0.1$).
      - The probability of a delay being caused by an accident is therefore 90% (i.e., $P(Accident) = 0.9$).
      - The likelihood of observing the delay given that it is caused by construction is 75% (i.e., $P(Delay|Construction) = 0.75$).
      - The likelihood of observing the delay given that it is caused by an accident is 25% (i.e., $P(Delay|Accident) = 0.25$).

      Given this, Mark uses Bayes' Theorem to update his belief about the likelihood of construction causing the delay. Derive the posterior probabilityof construction, given the observed delay.

   b) Mark has two choices:

      - If Mark takes the backroad, the travel time is fixed at 60 minutes and the utility is 0.
      - If Mark takes the highway:
        - If there is construction (posterior probability 0.1), the delay is 45 minutes, and the utility is -30.
        - If there is no construction (posterior probability 0.9), the travel time is 15 minutes, and the utility is 10.

      Calculate the expected utility for both taking the highway and taking the backroad, and determine the optimal decision for Mark.

4 Old exam 2022-05-24

1a) What are the differences between the Bayesian and classical methods?

1b) Explain how to perform diagnostic analysis in Bayesian models.

1c) What is a hierarchical model? Give an example of when this type of model is useful. What is the issue with the applicability of this type of model?

1d) What is a prior? Give examples of different types of priors and talk about their strengths and weaknesses.

2. Lupus is an autoimmune disease, where antibodies attack plasma proteins instead of foreign cells. It is believed that 2% of the population has this condition. Suppose that the exam to detect Lupus has 98% accuracy when the patient has the disease and 74% when (s)he does not have the condition.

2a) What theorem is involved in this type of problem?

2b) What is the probability that the patient has Lupus when the result is positive?

3. Suppose that if $\theta = 0$, then y follows a normal distribution with mean 3 and variance σ^2, and if $\theta = 1$, then y has a normal distribution with mean 4 and variance σ^2. Assuming that $P(\theta = 0) = 0.3$, $P(\theta = 1) = 0.7$, and $\sigma = 2$:

   3a) Write the formula for the marginal probability density for y.

   3b) What is $P(\theta = 1 \mid y = 1)$?

4. Hemophilia is a disease that exhibits X-chromosome-linked recessive inheritance that is fatal for women who inherit two such genes. Since human males have one X-chromosome and one Y-chromosome, whereas females have two X-chromosomes (one inherited from each parent), a female carrying the gene on only one of her two X-chromosomes is not affected.

   Consider a woman who has an affected brother, which implies that her mother must be a carrier of the hemophilia gene (with one "good" and one "bad" hemophilia gene). We are also told that her father is not affected; thus, the woman herself has a fifty-fifty chance of having the gene. The unknown quantity of interest, the state of the woman, has just two values: the woman is either a carrier of the gene ($\theta = 1$) or not ($\theta = 0$). Suppose she has three sons, none of whom are affected. Let $y_i = 1$ or 0 denote an affected or unaffected son, respectively. The outcomes of the three sons are exchangeable, and conditional on the unknown θ, are independent. We assume the sons are not identical triples.

   4a) What is the prior distribution for the unknown θ?

   4b) What is the likelihood function of the three independent data?

   4c) What is the posterior probability that the woman is a carrier?

   4d) Supposing that the woman has a fourth son, who is also unaffected, what is the posterior probability that the woman is a carrier?

5. Suppose y follows a binomial distribution for a given n and unknown parameter θ, where the prior distribution of θ is Beta(α, β). Find the marginal distribution of y.

6. A 45-year-old man has a tumor that is malignant with an 80% probability. Based on statistics:

   - Expected lifetime is 30 years if no cancer.
   - Expected lifetime is 10 years if cancer and radiation therapy are used.
   - Expected lifetime is 25 years if cancer and surgery, but the probability of dying in surgery is 50% (risky surgery).
   - Expected lifetime is 4 years if cancer and no treatment.

   Which treatment should be chosen to aim at having a quality-adjusted lifetime? Note that 3 years are subtracted from the time spent in treatments.

5 Book exercises

The following exercises are from the MS2505 course book[2]. Follow the cite in each question for the pages the exercise is at in the book.

1.1 Conditional probability: suppose that if $\theta = 1$, then y has a normal distribution with mean 1 and standard deviation α, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation α. Also, suppose $\Pr(\theta = 1) = 0.5$ and $\Pr(\theta = 2) = 0.5$[3].

  $a)$ For $\alpha = 2$, write the formula for the marginal probability density for y and sketch it.

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

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

1.4 We will use the football dataset to estimate some conditional probabilities about professional football games. There were twelve games with point spreads of 8 points; the outcomes in those games were: -7, -5, -3, -3, 1, 6, 7, 13, 15, 16, 20 and 21, with positive values indicating wins by the favorite and negative values indicating wins by the underdog. Consider the following conditional probabilities[3]:

- $\Pr(\text{favorite wins} \mid \text{point spread } = 8)$,
- $\Pr(\text{favorite wins by at least } 8 \mid \text{point spread } = 8)$,
- $\Pr(\text{favorite wins by at least } 8 \mid \text{point spread } = 8 \text{ and favorite wins})$.

1.6 Approximately $\frac{1}{125}$ of all births are fraternal twins and $\frac{1}{300}$ 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 $\frac{1}{2}$)[3].

2.1 Suppose you have a $Beta(4, 4)$ 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[4].

2.4 Let y be the number of 6's in 1000 independent rolls of a particular real die, which may be unfair. Let θ be the probability that the die lands on *6*. Suppose your prior distribution for θ is as follows[4]:

- $\Pr(\theta = 1/12) = 0.25$,
- $\Pr(\theta = 1/6) = 0.5$,
- $\Pr(\theta = 1/4) = 0.25$.

Using the normal approximation for the conditional distributions, $p(y|\theta)$, sketch your approximate prior predictive distribution for y.

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[4].

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

$$\Pr(y = k) = \int_0^1 \Pr(y = k|\theta)d\theta, k = 0, 1, ..., n.$$

$b)$ Suppose you assign a $Beta(\alpha, \beta)$ 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, $\frac{y}{n}$.

2.8 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 $y = 150$ 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[4].

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

$b)$ 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 $\tilde{y}$. (Your answer will still be a function of n.)

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

$d)$ Do the same for $n = 100$.

2.16 Suppose y has a binomial distribution for a given n and unknown parameter θ, where the prior distribution of θ is $Beta(\alpha, \beta)$. Find $p(y)$, the marginal distribution of y, for $y = 0, \ldots, n$ (unconditional on θ). This discrete distribution is known as the beta-binomial, for obvious reasons[4].

3.3 An experiment was performed on the effects of magnetic fields on the flow of calcium out of chicken brains. Two groups of chickens were involved: a control group of 32 chickens and an exposed group of 36 chickens. One measurement was taken on each chicken, and the purpose of the experiment was to measure the average flow μ_c in untreated (control) chickens and the average flow µt in treated chickens. The 32 measurements on the control group had a sample mean of 1.013 and a sample standard deviation of 0.24. The 36 measurements on the treatment group had a sample mean of 1.173 and a sample standard deviation of 0.20.

Assuming the control measurements were taken at random from a normal distribution with mean μ_c and variance σ_c^2, what is the posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of μ_t. Assume a uniform prior distribution on (μ_c, μ_t, $\log(\sigma_c)$, $\log(\sigma_t)$)[5].

3.5 Rounded data: it is a common problem for measurements to be observed in rounded form (for a review, see Heitjan, 1989). For a simple example, suppose we weigh an object five times and measure weights, rounded to the nearest pound, of 10, 10, 12, 11, 9. Assume the unrounded measurements are normally distributed with a noninformative prior distribution on the mean μ and variance σ^2[5].

$a)$ Give the posterior distribution for (μ, σ^2) obtained by pretending that the observations are exact unrounded measurements.

$b)$ Give the correct posterior distribution for (μ, σ^2) treating the measurements as rounded.

3.9 Suppose y is an independent and identically distributed sample of size n from the distribution $N(\mu, \sigma^2)$, where the prior distribution for (μ, σ^2) is $N-Inv-\chi^2(\mu, \sigma^2|\mu_0, \frac{\sigma_0^2}{\kappa_0}; v_0, \sigma_0^2)$; that is, $\sigma^2\ Inv-\chi^2(v_0, \sigma_0^2)$ and $\mu|\sigma^2\ N(\mu_0, \sigma^2/\kappa_0)$. The posterior distribution, $p(\mu, \sigma^2|y)$, is also normal-inverse-χ^2; derive explicitly its parameters in terms of the prior parameters and the sufficient statistics of the data[5].

4.2 Derive the analytic form of the information matrix and the normal approximation variance for the bioassay example[6][7].

4.4 Asymptotic normality: assuming the regularity conditions hold, we know that $p(\theta|y)$ approaches normality as $n \to \infty$. In addition, if $\Phi = f(\theta)$ is any one-to-one continuous transformation of θ, we can express the Bayesian inference in terms of Φ and find that $p(\Phi|y)$ also approaches normality. But a nonlinear transformation of a normal distribution is no longer normal. How can both limiting normal distributions be valid[7]?

4.13 Discuss the criticism, "Bayesianism assumes: (a) Either a weak or uniform prior [distribution], in which case why bother?, (b) Or a strong prior [distribution], in which case why collect new data?, (c) Or more realistically, something in between, in which case Bayesianism always seems to duck the issue" (Ehrenberg, 1986). Feel free to use any of the examples covered so far to illustrate your points.

5.4 Exchangeable prior distributions: suppose it is known a priori that the 2J parameters $\theta_1, \ldots, \theta_{2J}$ are clustered into two groups, with exactly half being drawn from a $N(1, 1)$ distribution, and the other half being drawn from a $N(-1, 1)$ distribution, but we have not observed which parameters come from which distribution.

Are $\theta_1, \ldots, \theta_{2J}$ exchangeable under this prior distribution?

5.7 If $y|\theta\ Poisson(\theta)$, and $\theta\ Gamma(\alpha, \beta)$, then the marginal (prior predictive) distribution of y is negative binomial with parameters α and β (or $p = \beta/(1 + \beta)$). Use the formulas (2.7)[1] and (2.8)[2] to derive the mean and variance of the negative binomial.

5.11 Suppose that in the rat tumor example, we wish to use a normal population distribution on the log-odds scale:

$$logit(\theta_j)\ N(\mu, \tau^2), for j = 1, \ldots, J.$$

As in Section 5.3, you will assign a noninformative prior distribution to the hyperparameters and perform a full Bayesian analysis.

Write the joint posterior density, $p(\theta, \mu, \tau|y)$.

[1] $E(\theta) = E(E(\theta|y))$

[2] $var(\theta) = E(var(\theta|y)) + var(E(\theta|y))$

References

[1] Bruna Palm. Ms2505: Bayesian statistics. Online, LP2 2024. Statistics course at Blekinge Institute of Technology.

[2] John B. Carlin, Davib B. Dunson, Andrew Gelman, Donald Rubin, Hal S. Ster, and Aki Vehtari. *Bayesian Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 6000 Broken Sound Parkway, NW, Boca Raton, US, 3rd edition, 11 2013. Free PDF version for non-profit purposes: https://users.aalto.fi/~ave/BDA3.pdf.

[3] John B. Carlin, David B. Dunson, Andrew Gelman, Donald Rubin, Hal S. Ster, and Aki Vehtari. *Bayesian Data Analysis*, chapter Chapter 1 Exercises, pages 27–28. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 6000 Broken Sound Parkway, NW, Boca Raton, US, 3rd edition, 2013.

[4] John B. Carlin, David B. Dunson, Andrew Gelman, Donald Rubin, Hal S. Ster, and Aki Vehtari. *Bayesian Data Analysis*, chapter Chapter 1 Exercises, pages 57–60. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 6000 Broken Sound Parkway, NW, Boca Raton, US, 3rd edition, 2013.

[5] John B. Carlin, David B. Dunson, Andrew Gelman, Donald Rubin, Hal S. Ster, and Aki Vehtari. *Bayesian Data Analysis*, chapter Chapter 1 Exercises, pages 79–80. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 6000 Broken Sound Parkway, NW, Boca Raton, US, 3rd edition, 2013.

[6] John B. Carlin, David B. Dunson, Andrew Gelman, Donald Rubin, Hal S. Ster, and Aki Vehtari. *Bayesian Data Analysis*, chapter Chapter 1 Exercises, pages 74–78. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 6000 Broken Sound Parkway, NW, Boca Raton, US, 3rd edition, 2013.

[7] John B. Carlin, David B. Dunson, Andrew Gelman, Donald Rubin, Hal S. Ster, and Aki Vehtari. *Bayesian Data Analysis*, chapter Chapter 1 Exercises, pages 98–100. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 6000 Broken Sound Parkway, NW, Boca Raton, US, 3rd edition, 2013.

Blekinge Institute of Technology
Institute of mathematics
and natural sciences

Tables and Formulas

Mathematical Statistics

(Intended for use at written examination)

by

Mattias Dahl, Efraim Laksman och Robert Nyqvist

1 Distributions

1.1 Continuous distributions (density functions)

1.1.1 Uniform distribution

$$f_X(x) = \frac{1}{b-a}, a \leq x \leq b, \quad E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}$$

1.1.2 Exponential distribution

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0, \quad E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}$$

1.1.3 Weibull distribution

$$f_X(x) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} e^{-(x/b)^c}, x \geq 0, \quad E(X) = b\Gamma((c+1)/c)$$

$$V(X) = b^2(\Gamma((c+2)/c) - \Gamma((c+2)/c)^2), \text{ Gammafunction} : \Gamma(x) = \int_0^\infty \mathrm{e}^{-y} y^{x-1} dy$$

1.1.4 Gaussian distribution (normal distribution)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad E(X) = m, \quad V(X) = \sigma^2$$

1.2 Discrete distributions (probability functions)

1.2.1 Binomial distribution

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n, \quad E(X) = np, \quad V(X) = np(1-p)$$

1.2.2 Hypergeometric distribution

$$p_X(x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}, \quad E(X) = np, \quad V(X) = npq\frac{N-n}{N-1}$$

1.2.3 Poisson distribution

$$p_X(x) = e^{-m}\frac{m^x}{x!}, x = 0, 1, 2, \ldots, \quad E(X) = V(X) = m$$

1.3 Approximations

1.3.1 Binomial distribution may be approximated with Poisson distribution if $p \leq 0.1, n \geq 10$.

1.3.2 Binomial distribution may be approximated with normal distribution if $npq \geq 10$.

1.3.3 Poisson distribution may be approximated with normal distribution if $m \geq 15$.

1.4 Gauss formulas for approximation:

$$E(g(X)) \approx g(E(X)),$$
$$V(g(X)) \approx (g'(E(X)))^2 V(X)$$

2 Stochastic processes

2.1 Stochastic processes in discrete time

2.1.1 Stationary (and asymptotic) state

$$\pi = \pi P$$

2.2 Stochastic processes in continuous time

2.2.1 Connection between hazard function $\lambda(t)$ and reliability $R(t)$

$$R(t) = e^{-\int_0^t \lambda(u)du}$$
$$\lambda(t) = -\frac{R'(t)}{R(t)}$$

2.2.2 The distribution vector $p(t)$ and intensity matrix A satisfy the relation

$$p'(t) = p(t)A$$

2.2.3 Stationary (and asymptotic) state

$$0 = \pi A$$

2.2.4 Erlang's formula

$$\pi_k = \frac{\rho^k/k!}{\rho^0/0! + \rho^1/1! + \rho^2/2! + \ldots + \rho^n/n!}, k = 0, 1, \ldots, n$$

3 Estimations

3.1 Point estimations

3.1.1 Expected value

$$\hat{m} = \bar{x} = \frac{1}{n}\sum x$$

3.1.2 Standard deviation

$$s = \sqrt{\frac{1}{n-1}(\sum x^2 - \frac{1}{n}(\sum x)^2)}$$

3.1.3 The least square function

$$Q_0(x_1, \ldots, x_n|\theta) = \sum(x - m(\theta))^2$$

3.1.4 The likelihood function (continuous s.v.)

$$L(x_1, \ldots, x_n|\theta) = f_{X_1}(x_1|\theta) \cdots f_{X_n}(x_n|\theta)$$

3.1.5 The likelihood function (discrete s.v.)

$$L(x_1, \ldots, x_n|\theta) = p_{X_1}(x_1|\theta) \cdots p_{X_n}(x_n|\theta)$$

3.2 Confidence intervals

3.2.1 Normal distribution with known standard deviation

$$\bar{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

3.2.2 Normal distribution with unknown standard deviation

$$\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

3.2.3 Prediction interval with linear regression

$$\hat{a} + \hat{b}x_0 \pm t(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

3.2.4 Comparison of expected values in two groups with normal distribution with known standard deviation

$$\bar{x} - \bar{y} \pm \lambda_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

3.2.5 Comparison of expected value in two groups with normal distribution with unknown standard deviation

$$\bar{x} - \bar{y} \pm t_{\alpha/2}(n_x - 1 + n_y - 1)s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}},$$

where

$$s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x - 1 + n_y - 1}}$$

3.2.6 Normal approximation of binomial distribution

$$\hat{p} \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

3.2.7 Normal approximation of Poisson distribution

$$\bar{x} \pm \lambda_{\alpha/2}\sqrt{\frac{\bar{x}}{n}}$$

3.2.8 The standard deviation in a normal distribution

$$\left(\sqrt{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}}, \sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}}\right)$$

4 Hypothesis testing

4.1 Test variables

4.1.1 Normal distribution with known standard deviation

$$u = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}}$$

4.1.2 Normal distribution with unknown standard deviation

$$u = \frac{\bar{x} - m_0}{s/\sqrt{n}}$$

4.1.3 Comparison of expected value in two groups of normal distribution with known standard deviation

$$u = \frac{\bar{x} - \bar{y} - C}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

4.1.4 Comparison of expected value in two groups of normal distribution with unknown but similar standard deviation

$$u = \frac{\bar{x} - \bar{y} - C}{s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}},$$

where

$$s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x - 1 + n_y - 1}}$$

4.1.5 Normal approximation of binomial distribution

$$u = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

4.1.6 Normal approximation of Poisson distribution

$$u = \frac{\bar{x} - m_0}{\sqrt{m_0/n}}$$

4.1.7 Comparison with some distribution

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

with $k - a - 1$ degrees of freedom.

4.1.8 Comparison with other observations

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

with $(r-1)(k-1)$ degrees of freedom.

4.2 Simple linear regression

4.2.1 Model

$$y = a + bx + \epsilon$$

4.2.2 Sums of squares

$$\begin{aligned} S_{xx} &= \sum x^2 - \frac{1}{n}(\sum x)^2 = \sum (x - \bar{x})^2 \\ S_{xy} &= \sum xy - \frac{1}{n}(\sum x)(\sum y) = \sum (x - \bar{x})(y - \bar{y}) \\ S_{yy} &= \sum y^2 - \frac{1}{n}(\sum y)^2 = \sum (y - \bar{y})^2 \end{aligned}$$

4.2.3 Estimators of parameters

$$\hat{b} = \frac{S_{xy}}{S_{xx}}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

4.2.4 Sum of squares errors

$$Q_0 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

4.2.5 Estimator of standard deviation of residuals

$$s = \sqrt{\frac{Q_0}{n-2}}$$

4.2.6 Estimators of standard deviation of estimators of parameters

$$s_a = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}, \quad s_b = \frac{s}{\sqrt{S_{xx}}}$$

4.2.7 Hypothesis test for parameters

$$u = \frac{a^* - a_0}{s_a}, \quad u = \frac{b^* - b_0}{s_b}$$

4.2.8 Confidence interval for simple linear regression

$$\hat{a} + \hat{b}x_0 \pm t(n-2)s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

4.2.9 Correlations

$$r_{Pearson} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$
$$r_{Spearman} = 1 - \frac{6\sum d^2}{n^3 - n}$$

4.2.10 Hypothesis test for correlation

$$u = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

4.2.11 Prediction interval for simple linear regression

$$\hat{a} + \hat{b}x_0 \pm t(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

Summary of main results

Bayesiansk statistik
MS2505

Bruna G. Palm

1. Beta(α, β) distribution:

$$\begin{aligned} f_Y(y) =& \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-\theta)^{\beta-1} \\ \mathrm{E}(Y) =& \frac{\alpha}{\alpha+\beta} \\ var(Y) =& \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned}$$

2. mean and variance of conditional distributions

$$\begin{aligned} \mathrm{E}(u) =& \mathrm{E}(\mathrm{E}(u|v)) = \int\int up(u,v)dudv = \int\int up(u|v)dup(v) = \int \mathrm{E}(u|v)p(v)dv \\ var(u) =& 4\,\mathrm{E}(var(u|v)) + var(\mathrm{E}(u|v)) \end{aligned}$$

3. Bayes' rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

4. The unnormalized posterior density:

$$p(\theta|y) \propto p(\theta)p(y|\theta)$$

5. Normalization term Z (constant given y):

$$Z = \int_0^1 \theta^y(1-\theta)^{n-y}d\theta = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

6. Binomial with Uniform prior
   - Posterior is

     $$p(\theta|y,n,M) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y},$$

     which is called Beta distribution

     $$\theta|y,n \sim \mathrm{Beta}(y+1, n-y+1)$$

   - Expected value

     $$\mathrm{E}[\theta|y] = \frac{y+1}{n+2}$$

   - Prior predictive distribution for new $\tilde{y}$ (discrete)

     $$p(\tilde{y}=1|M) = \int_0^1 p(\tilde{y}=1|\theta,y,n,M)p(\theta|M)d\theta$$

   - Posterior predictive distribution for new $\tilde{y}$ (discrete)

     $$p(\tilde{y}=1|y,n,M) = \int_0^1 p(\tilde{y}=1|\theta,y,n,M)p(\theta|y,n,M)d\theta$$

7. Binomial with Beta prior
   - Prior

     $$\mathrm{Beta}(\theta|\alpha,\beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- Posterior

$$\begin{aligned} p(\theta|y,n,M) &\propto \theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \end{aligned}$$

after normalization

$$p(\theta|y,n,M) = \text{Beta}(\theta|\alpha+y,\beta+n-y)$$

- Posterior mean

$$\mathrm{E}[\theta|y] = \frac{\alpha+y}{\alpha+\beta+n}$$

- Posterior variance

$$\mathrm{Var}[\theta|y] = \frac{\mathrm{E}[\theta|y](1-\mathrm{E}[\theta|y])}{\alpha+\beta+n+1}$$

8. Normal distribution with conjugate prior (Normal) for θ

- Assume σ^2 known
  - Likelihood

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

  - Prior

$$p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right)$$

  - Posterior

$$\begin{aligned} p(\theta|y) &\propto \exp\left(-\frac{1}{2}\left[\frac{(y-\theta)^2}{\sigma^2}+\frac{(\theta-\mu_0)^2}{\tau_0^2}\right]\right) \\ &\propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right) \\ \theta|y &\sim \mathrm{N}(\mu_1,\tau_1^2), \quad \text{where} \quad \mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0+\frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2}+\frac{1}{\sigma^2} \end{aligned}$$

  - Several observations $y = (y_1, \ldots, y_n)$

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

$$\text{where} \quad \mu_n = \frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2}+\frac{n}{\sigma^2}$$

  - Posterior predictive distribution

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\ p(\tilde{y}|y) &\propto \int \exp\left(-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2\right)\exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right)d\theta \end{aligned}$$

$$\tilde{y}|y \sim \mathrm{N}(\mu_1,\sigma^2+\tau_1^2), \quad \text{where} \quad \tau_1^2 \quad \text{is the posterior variance}$$

9. Monte Carlo and posterior draws

$$E_{p(\theta|y)}[\theta] = \int \theta p(\theta \mid y) \approx \frac{1}{S}\sum_{s=1}^{S}\theta^{(s)}$$

10. Joint distribution of parameters

$$p(\theta_1,\theta_2 \mid y) \propto p(y \mid \theta_1,\theta_2)p(\theta_1,\theta_2)$$

11. Marginalization

$$p(\theta_1 \mid y) = \int p(\theta_1,\theta_2 \mid y)d\theta_2$$

$p(\theta_1 \mid y)$ is a marginal distribution

12. Monte Carlo approximation

$$p(\theta_1 \mid y) \approx \frac{1}{S}\sum_{s=1}^{S}p(\theta_1,\theta_2^{(s)} \mid y),$$

where $\theta_2^{(s)}$ are draws from $p(\theta_2 \mid y)$

13. Marginalization over posterior distribution

$$
\begin{aligned}
p(\tilde{y} \mid y) &= \int p(\tilde{y} \mid \theta) p(\theta \mid y) d\theta \\
&= \int p(\tilde{y}, \theta \mid y) d\theta
\end{aligned}
$$

$p(\tilde{y} \mid y)$ is a predictive distribution

14. Gaussian non-informative prior (Uniform)

$$
\begin{aligned}
p(\mu, \sigma^2) &\propto \sigma^{-2} \\
p(\mu, \sigma^2 \mid y) &\propto \sigma^{-2} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\
p(\mu, \sigma^2 \mid y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right)
\end{aligned}
$$

- Marginal posterior $p(\sigma^2 \mid y)$
  - Known mean

$$
\begin{aligned}
\sigma^2 \mid y &\sim \mathrm{Inv} - \chi^2(n, v) \\
\text{where} \quad v &= \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta)^2
\end{aligned}
$$

  - Unknown mean

$$
\begin{aligned}
\sigma^2 \mid y &\sim \mathrm{Inv} - \chi^2(n-1, s^2) \\
\text{where} \quad s^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2
\end{aligned}
$$

- Marginal posterior $p(\mu \mid y)$

$$
\begin{aligned}
p(\mu \mid y) &\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2} \\
p(\mu \mid y) &= t_{n-1}(\mu \mid \bar{y}, s^2/n)
\end{aligned}
$$

- Posterior predictive distribution given known variance

$$
\begin{aligned}
p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\
&= \int \mathrm{N}(\tilde{y} \mid \mu, \sigma^2)\, \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) d\mu \\
&= \mathrm{N}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})\sigma^2)
\end{aligned}
$$

- Posterior predictive distribution given unknown variance

$$
p(\tilde{y} \mid y) = t_{n-1}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})s^2)
$$

15. Gaussian conjugate prior

- Conjugate prior has to have a form $p(\sigma^2)p(\mu \mid \sigma^2)$

$$
\begin{aligned}
\mu \mid \sigma^2 &\sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0) \\
\sigma^2 &\sim \mathrm{Inv} - \chi^2(\nu_0, \sigma_0^2) \\
p(\mu, \sigma^2) &= \mathrm{N\,Inv} - \chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)
\end{aligned}
$$

- Joint posterior

$$
\begin{aligned}
p(\mu, \sigma^2 \mid y) \approx\, & \sigma^{-1}(\sigma^2)^{(-\nu_0/2+1)} \exp\{\frac{-1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2]\} \\
& \times (\sigma^2)^{-n/2} \exp\{\frac{-1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\} \\
p(\mu, \sigma^2 \mid y) =\, & \mathrm{N\,Inv} - \chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2)
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n \sigma_n^2 &= \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2
\end{aligned}
$$

16. Normal approximation (large sample) Normal approximation

$$p(\theta|y) \approx \mathrm{N}(\hat{\theta}, [I(\hat{\theta})]^{-1})$$

where

$$I(\theta) = -\frac{d^2}{d\theta^2} \log p(\theta|y)$$

17. Posterior odds

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)}{p(\theta_2)p(y|\theta_2)}$$