Задача 1. Рассматривается модель линейной регрессии $Y_i = \mu + \varepsilon_i$, $i = 1, \ldots, n$, где случайные ошибки $\varepsilon_1, \ldots, \varepsilon_n$ независимы и имеют стандартное нормальное распределение.

(a) Найдите апостериорное распределение параметра μ, считая, что априорное распределение параметра μ есть $N(0,1)$.

(b) Найдите байесовскую точечную оценку $\widehat{\mu}$ как математическое ожидание апостериорного распределения параметра μ.

(c) Постройте байесовский 95%-ый доверительный интервал для параметра μ.

(d) Рассчитайте числовые значения точечной и интервальной оценок параметра μ, если известно, что $n = 2$, $y_1 = 0$, $y_2 = 1.5$.

(e) Используя данные из предыдущего пункта, протестируйте гипотезу $H_0 : \mu < 1$ против альтернативной гипотезы $H_1 : \mu \geq 1$.

Решение. (a) Имеем:

$$\varphi(\mu \mid y) \propto \mathcal{L}(y \mid \mu) \cdot \pi(\mu) = \prod_{i=1}^{n} \left[(2\pi)^{-1/2} e^{-\frac{(y_i-\mu)^2}{2}} \right] \cdot (2\pi)^{-1/2} e^{-\frac{\mu^2}{2}} \propto$$

$$\propto e^{-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2} - \frac{\mu^2}{2}} = e^{-\frac{\sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} y_i \cdot \mu + n\mu^2 + \mu^2}{2}} \propto e^{-\frac{(n+1)\mu^2 - 2\sum_{i=1}^{n} y_i \cdot \mu}{2}} =$$

$$= e^{-\frac{(n+1)\left[\mu^2 - 2\mu \frac{\sum_{i=1}^{n} y_i}{n+1}\right]}{2}} = e^{-\frac{(n+1)\left[\mu^2 - 2\mu \frac{\sum_{i=1}^{n} y_i}{n+1} + \left(\frac{\sum_{i=1}^{n} y_i}{n+1}\right)^2\right] - (n+1)\left(\frac{\sum_{i=1}^{n} y_i}{n+1}\right)^2}{2}} \propto$$

$$\propto e^{-\frac{(n+1)\left[\mu - \frac{\sum_{i=1}^{n} y_i}{n+1}\right]^2}{2}} = e^{-\frac{\left[\mu - \frac{\sum_{i=1}^{n} y_i}{n+1}\right]^2}{2 \cdot \frac{1}{n+1}}},$$

т.е. $(\mu \mid Y = y) \sim N\left(\frac{\sum_{i=1}^{n} y_i}{n+1}, \frac{1}{n+1} \right)$ — апостериорное распределение параметра μ.

(b) Поскольку математическое ожидание апостериорного распределения параметра μ есть $\frac{\sum_{i=1}^{n} y_i}{n+1}$, то требуемая точечная оценка $\widehat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n+1}$.

(c) Поскольку $(\mu \mid Y = y) \sim N\left(\frac{\sum_{i=1}^{n} y_i}{n+1}, \frac{1}{n+1} \right)$, то при условии $Y = y$ с вероятностью 95% имеет место неравенство: $z_{0.025} \leq \dfrac{\mu - \frac{\sum_{i=1}^{n} y_i}{n+1}}{\sqrt{\frac{1}{n+1}}} \leq z_{0.975}$, где точки $z_{0.025}$ и $z_{0.975}$ означают квантили стандартного нормального распределения уровней 0.025 и 0.975 соответственно. Значит, байесовская интервальная оценка параметра μ есть:

$$\frac{\sum_{i=1}^{n} y_i}{n+1} + z_{0.025}\sqrt{\frac{1}{n+1}} \leq \mu \leq \frac{\sum_{i=1}^{n} y_i}{n+1} + z_{0.975}\sqrt{\frac{1}{n+1}}.$$

(d) Числовое значение точечной байесовской оценки параметра μ есть:

$$\widehat{\mu} = \frac{0 + 1.5}{3} = 0.5.$$

Числовое значение интервальной байесовской оценки параметра μ есть:

$$\underbrace{0.5-1.96\sqrt{\frac{1}{3}}}_{=-0.63}\le\mu\le\underbrace{0.5+1.96\sqrt{\frac{1}{3}}}_{=1.63}.$$

(e) Имеем:

$$\mathbb{P}(\{\mu<1\}\,|\,\{Y=y\})=\mathbb{P}\left(\left\{\frac{\mu-0.5}{\sqrt{1/3}}<\frac{1-0.5}{\sqrt{1/3}}\right\}\,|\,\{Y=y\}\right)=\Phi\left(\frac{1-0.5}{\sqrt{1/3}}\right)=$$

$$=\Phi\left(0.87\right)=\text{normcdf}(0.87,0,1)=0.81,$$

$$\mathbb{P}(\{\mu\ge1\}\,|\,\{Y=y\})=1-\mathbb{P}(\{\mu<1\}\,|\,\{Y=y\})=1-0.81=0.19.$$

Поскольку вероятность $\mathbb{P}(\{\mu<1\}\,|\,\{Y=y\})$ больше вероятности $\mathbb{P}(\{\mu\ge1\}\,|\,\{Y=y\})$, то должна быть принята гипотеза H_0. ☐

Задача 2. Рассматривается модель линейной регрессии $Y_i=\mu+\varepsilon_i$, $i=1,\ldots,n$, где случайные ошибки $\varepsilon_1,\ldots,\varepsilon_n$ независимы и имеют стандартное нормальное распределение.

(a) Найдите апостериорное распределение параметра μ, считая, что априорное распределение параметра μ есть $N(2,1)$.

(b) Найдите байесовскую точечную оценку $\widehat{\mu}$ как математическое ожидание апостериорного распределения параметра μ.

(c) Постройте байесовский 95%-ый доверительный интервал для параметра μ.

(d) Рассчитайте числовые значения точечной и интервальной оценок параметра μ, если известно, что $n=2$, $y_1=1$, $y_2=3$.

(e) Используя данные из предыдущего пункта, протестируйте гипотезу $H_0:\mu<1$ против альтернативной гипотезы $H_1:\mu\ge1$.

Решение. (a) Имеем:

$$\varphi(\mu\,|\,y)\propto\mathcal{L}(y\,|\,\mu)\cdot\pi(\mu)=\prod\nolimits_{i=1}^{n}\left[(2\pi)^{-1/2}e^{-\frac{(y_i-\mu)^2}{2}}\right]\cdot(2\pi)^{-1/2}e^{-\frac{(\mu-2)^2}{2}}\propto$$

$$\propto e^{-\frac{\sum_{i=1}^{n}(y_i-\mu)^2+(\mu-2)^2}{2}}=e^{-\frac{\sum_{i=1}^{n}y_i^2-2\mu\sum_{i=1}^{n}y_i+n\mu^2+\mu^2-4\mu+4}{2}}\propto e^{-\frac{(n+1)\mu^2-2\mu\left(\sum_{i=1}^{n}y_i+2\right)}{2}}=$$

$$=e^{-\frac{(n+1)\left[\mu^2-2\mu\frac{\sum_{i=1}^{n}y_i+2}{n+1}\right]}{2}}=e^{-\frac{(n+1)\left[\mu^2-2\mu\frac{\sum_{i=1}^{n}y_i+2}{n+1}+\left(\frac{\sum_{i=1}^{n}y_i+2}{n+1}\right)^2\right]-(n+1)\left(\frac{\sum_{i=1}^{n}y_i+2}{n+1}\right)^2}{2}}\propto$$

$$\propto e^{-\frac{(n+1)\left[\mu-\frac{\sum_{i=1}^{n}y_i+2}{n+1}\right]^2}{2}}=e^{-\frac{\left[\mu-\frac{\sum_{i=1}^{n}y_i+2}{n+1}\right]^2}{2\cdot\frac{1}{n+1}}},$$

т.е. $(\mu\,|\,Y=y)\sim N\left(\dfrac{\sum_{i=1}^{n}y_i+2}{n+1},\dfrac{1}{n+1}\right)$ — апостериорное распределение параметра μ.

(b) Поскольку математическое ожидание апостериорного распределения параметра μ есть $\dfrac{\sum_{i=1}^{n}y_i+2}{n+1}$, то требуемая точечная оценка $\widehat{\mu}=\dfrac{\sum_{i=1}^{n}y_i+2}{n+1}$.

(c) Поскольку $(\mu \mid Y = y) \sim N\left(\frac{\sum_{i=1}^{n} y_i + 2}{n+1}, \frac{1}{n+1}\right)$, то при условии $Y = y$ с вероятностью 95% имеет место неравенство: $z_{0.025} \le \frac{\mu - \frac{\sum_{i=1}^{n} y_i + 2}{n+1}}{\sqrt{\frac{1}{n+1}}} \le z_{0.975}$, где точки $z_{0.025}$ и $z_{0.975}$ означают квантили стандартного нормального распределения уровней 0.025 и 0.975 соответственно. Значит, байесовская интервальная оценка параметра μ есть:

$$\frac{\sum_{i=1}^{n} y_i + 2}{n+1} + z_{0.025}\sqrt{\frac{1}{n+1}} \le \mu \le \frac{\sum_{i=1}^{n} y_i + 2}{n+1} + z_{0.975}\sqrt{\frac{1}{n+1}}.$$

(d) Числовое значение точечной байесовской оценки параметра μ есть:

$$\hat{\mu} = \frac{1+3+2}{3} = 2.$$

Числовое значение интервальной байесовской оценки параметра μ есть:

$$\underbrace{2 - 1.96\sqrt{\frac{1}{3}}}_{=0.87} \le \mu \le \underbrace{2 + 1.96\sqrt{\frac{1}{3}}}_{=3.13}.$$

(e) Имеем:

$$\mathbb{P}(\{\mu < 1\} \mid \{Y = y\}) = \mathbb{P}\left(\left\{\frac{\mu - 2}{\sqrt{1/3}} < \frac{1-2}{\sqrt{1/3}}\right\} \mid \{Y = y\}\right) = \Phi\left(\frac{1-2}{\sqrt{1/3}}\right) =$$

$$= \Phi(-1.73) = \text{normcdf}(-1.73, 0, 1) = 0.04,$$

$$\mathbb{P}(\{\mu \ge 1\} \mid \{Y = y\}) = 1 - \mathbb{P}(\{\mu < 1\} \mid \{Y = y\}) = 1 - 0.04 = 0.96.$$

Поскольку вероятность $\mathbb{P}(\{\mu \ge 1\} \mid \{Y = y\})$ больше вероятности $\mathbb{P}(\{\mu < 1\} \mid \{Y = y\})$, то должна быть принята гипотеза H_1. □

Задача 3. Рассматривается модель линейной регрессии $Y_i = \beta x_i + \varepsilon_i$, $i = 1, \ldots, n$, где случайные ошибки $\varepsilon_1, \ldots, \varepsilon_n$ независимы и имеют стандартное нормальное распределение, а $x_1, \ldots, x_n$ — известные числа, не равные нулю одновременно.

(a) Найдите апостериорное распределение параметра β, считая, что априорное распределение параметра β есть $N(0, 1)$.

(b) Найдите байесовскую точечную оценку $\hat{\beta}$ как математическое ожидание апостериорного распределения параметра β.

(c) Постройте байесовский 95%-ый доверительный интервал для параметра β.

(d) Рассчитайте числовые значения точечной и интервальной оценок параметра β, если известно, что $n = 2$, $x_1 = 1$, $y_1 = 1$, $x_2 = 2$, $y_2 = 3$.

(e) Используя данные из предыдущего пункта, протестируйте гипотезу $H_0 : \beta < 1$ против альтернативной гипотезы $H_1 : \beta \ge 1$.

Решение. (a) Имеем:

$$\varphi(\beta \mid y) \propto \mathcal{L}(y \mid \beta) \cdot \pi(\beta) = \prod_{i=1}^{n}\left[(2\pi)^{-1/2} e^{-\frac{(y_i - \beta x_i)^2}{2}}\right] \cdot (2\pi)^{-1/2} e^{-\frac{\beta^2}{2}} \propto$$

$$\propto e^{-\frac{\sum_{i=1}^{n}(y_i - \beta x_i)^2}{2} - \frac{\beta^2}{2}} = e^{-\frac{\sum_{i=1}^{n} y_i^2 - 2\beta\sum_{i=1}^{n} x_i y_i + \beta^2 \sum_{i=1}^{n} x_i^2 + \beta^2}{2}} \propto e^{-\frac{\left(\sum_{i=1}^{n} x_i^2 + 1\right)\beta^2 - 2\beta\sum_{i=1}^{n} x_i y_i}{2}} =$$

$$= e^{-\frac{\left(\sum_{i=1}^n x_i^2+1\right)\left[\beta^2-2\beta\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}\right]}{2}} = e^{-\frac{\left(\sum_{i=1}^n x_i^2+1\right)\left[\beta^2-2\beta\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}+\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}\right)^2\right]-\left(\sum_{i=1}^n x_i^2+1\right)\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}\right)^2}{2}} \propto$$

$$\propto e^{-\frac{\left(\sum_{i=1}^n x_i^2+1\right)\left[\beta-\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}\right]^2}{2}} = e^{-\frac{\left[\beta-\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}\right]^2}{2\cdot\frac{1}{\sum_{i=1}^n x_i^2+1}}},$$

т.е. $(\beta \mid Y = y) \sim N\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}, \frac{1}{\sum_{i=1}^n x_i^2+1}\right)$ — апостериорное распределение параметра β.

(b) Поскольку математическое ожидание апостериорного распределения параметра β есть $\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}$, то требуемая точечная оценка $\widehat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}$.

(c) Поскольку $(\beta \mid Y = y) \sim N\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}, \frac{1}{\sum_{i=1}^n x_i^2+1}\right)$, то при условии $Y = y$ с вероятностью 95% имеет место неравенство: $z_{0.025} \le \frac{\beta - \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1}}{\sqrt{\frac{1}{\sum_{i=1}^n x_i^2+1}}} \le z_{0.975}$, где точки $z_{0.025}$ и $z_{0.975}$ означают квантили стандартного нормального распределения уровней 0.025 и 0.975 соответственно. Значит, байесовская интервальная оценка параметра β есть:

$$\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1} + z_{0.025}\sqrt{\frac{1}{\sum_{i=1}^n x_i^2+1}} \le \beta \le \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2+1} + z_{0.975}\sqrt{\frac{1}{\sum_{i=1}^n x_i^2+1}}.$$

(d) Числовое значение точечной байесовской оценки параметра β есть:

$$\widehat{\beta} = \frac{1\times1+2\times3}{1^2+2^2+1} = \frac{7}{6}.$$

Числовое значение интервальной байесовской оценки параметра μ есть:

$$\underbrace{\frac{7}{6}-1.96\sqrt{\frac{1}{6}}}_{=0.37} \le \mu \le \underbrace{\frac{7}{6}+1.96\sqrt{\frac{1}{6}}}_{=1.97}.$$

(e) Имеем:

$$\mathbb{P}(\{\beta<1\} \mid \{Y=y\}) = \mathbb{P}\left(\left\{\frac{\beta-7/6}{\sqrt{1/6}} < \frac{1-7/6}{\sqrt{1/6}}\right\} \mid \{Y=y\}\right) = \Phi\left(\frac{1-7/6}{\sqrt{1/6}}\right) =$$

$$= \Phi\left(-0.41\right) = \text{normcdf}(-0.41, 0, 1) = 0.34,$$

$$\mathbb{P}(\{\beta \ge 1\} \mid \{Y=y\}) = 1 - \mathbb{P}(\{\beta<1\} \mid \{Y=y\}) = 1-0.34 = 0.66.$$

Поскольку вероятность $\mathbb{P}(\{\beta \ge 1\} \mid \{Y=y\})$ больше вероятности $\mathbb{P}(\{\beta<1\} \mid \{Y=y\})$, то должна быть принята гипотеза H_1. ☐

Задача 4. Рассматривается модель линейной регрессии $Y_i = \beta x_i + \varepsilon_i$, $i = 1, \ldots, n$, где случайные ошибки $\varepsilon_1, \ldots, \varepsilon_n$ независимы и имеют стандартное нормальное распределение, а $x_1, \ldots, x_n$ — известные числа, не равные нулю одновременно.

(a) Найдите апостериорное распределение параметра β, считая, что априорное распределение параметра β есть $N(-2, 1)$.

(b) Найдите байесовскую точечную оценку $\widehat{\beta}$ как математическое ожидание апостериорного распределения параметра β.

(c) Постройте байесовский 95%-ый доверительный интервал для параметра β.

(d) Рассчитайте числовые значения точечной и интервальной оценок параметра β, если известно, что $n = 2$, $x_1 = 1$, $y_1 = -1$, $x_2 = 2$, $y_2 = -3$.

(e) Используя данные из предыдущего пункта, протестируйте гипотезу $H_0 : \beta < -1$ против альтернативной гипотезы $H_1 : \beta \geq -1$.

Решение. (a) Имеем:

$$\varphi(\beta \mid y) \propto \mathcal{L}(y \mid \beta) \cdot \pi(\beta) = \prod_{i=1}^{n} \left[(2\pi)^{-1/2} e^{-\frac{(y_i - \beta x_i)^2}{2}} \right] \cdot (2\pi)^{-1/2} e^{-\frac{(\beta+2)^2}{2}} \propto$$

$$\propto e^{-\frac{\sum_{i=1}^{n}(y_i - \beta x_i)^2 + (\beta+2)^2}{2}} = e^{-\frac{\sum_{i=1}^{n} y_i^2 - 2\beta \sum_{i=1}^{n} x_i y_i + \beta^2 \sum_{i=1}^{n} x_i^2 + \beta^2 + 4\beta + 4}{2}} \propto e^{-\frac{\left(\sum_{i=1}^{n} x_i^2 + 1\right)\beta^2 - 2\beta\left(\sum_{i=1}^{n} x_i y_i - 2\right)}{2}} =$$

$$= e^{-\frac{\left(\sum_{i=1}^{n} x_i^2 + 1\right)\left[\beta^2 - 2\beta \frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}\right]}{2}} = e^{-\frac{\left(\sum_{i=1}^{n} x_i^2 + 1\right)\left[\beta^2 - 2\beta \frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1} + \left(\frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}\right)^2\right] - \left(\sum_{i=1}^{n} x_i^2 + 1\right)\left(\frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}\right)^2}{2}} \propto$$

$$\propto e^{-\frac{\left(\sum_{i=1}^{n} x_i^2 + 1\right)\left[\beta - \frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}\right]^2}{2}} = e^{-\frac{\left[\beta - \frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}\right]^2}{2 \cdot \frac{1}{\sum_{i=1}^{n} x_i^2 + 1}}},$$

т.е. $(\beta \mid Y = y) \sim N\left(\frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}, \frac{1}{\sum_{i=1}^{n} x_i^2 + 1} \right)$ — апостериорное распределение параметра β.

(b) Поскольку математическое ожидание апостериорного распределения параметра β есть $\frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}$, то требуемая точечная оценка $\widehat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}$.

(c) Поскольку $(\beta \mid Y = y) \sim N\left(\frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}, \frac{1}{\sum_{i=1}^{n} x_i^2 + 1} \right)$, то при условии $Y = y$ с вероятностью 95% имеет место неравенство: $z_{0.025} \leq \dfrac{\beta - \frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1}}{\sqrt{\frac{1}{\sum_{i=1}^{n} x_i^2 + 1}}} \leq z_{0.975}$, где точки $z_{0.025}$ и $z_{0.975}$ означают квантили стандартного нормального распределения уровней 0.025 и 0.975 соответственно. Значит, байесовская интервальная оценка параметра β есть:

$$\frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1} + z_{0.025}\sqrt{\frac{1}{\sum_{i=1}^{n} x_i^2 + 1}} \le \beta \le \frac{\sum_{i=1}^{n} x_i y_i - 2}{\sum_{i=1}^{n} x_i^2 + 1} + z_{0.975}\sqrt{\frac{1}{\sum_{i=1}^{n} x_i^2 + 1}} .$$

(d) Числовое значение точечной байесовской оценки параметра β есть:

$$\widehat{\beta} = \frac{1\times(-1) + 2\times(-3) - 2}{1^2 + 2^2 + 1} = -\frac{9}{6} = -\frac{3}{2} .$$

Числовое значение интервальной байесовской оценки параметра μ есть:

$$\underbrace{-\frac{3}{2} - 1.96\sqrt{\frac{1}{6}}}_{=-2.30} \le \mu \le \underbrace{-\frac{3}{2} + 1.96\sqrt{\frac{1}{6}}}_{=-0.7} .$$

(e) Имеем:

$$\mathbb{P}(\{\beta < -1\} \mid \{Y = y\}) = \mathbb{P}\left(\left\{\frac{\beta + 3/2}{\sqrt{1/6}} < \frac{-1 + 3/2}{\sqrt{1/6}}\right\} \mid \{Y = y\}\right) = \Phi\left(\frac{-1 + 3/2}{\sqrt{1/6}}\right) =$$

$$= \Phi\left(1.22\right) = \text{normcdf}(1.22, 0, 1) = 0.89 ,$$

$$\mathbb{P}(\{\beta \ge -1\} \mid \{Y = y\}) = 1 - \mathbb{P}(\{\beta < -1\} \mid \{Y = y\}) = 1 - 0.89 = 0.11 .$$

Поскольку вероятность $\mathbb{P}(\{\beta < -1\} \mid \{Y = y\})$ больше вероятности $\mathbb{P}(\{\beta \ge -1\} \mid \{Y = y\})$, то должна быть принята гипотеза H_0. □