Formula sheet for Bayesian Learning

Course codes: ST5301 and ST8101

Stockholms universitet

Version 1.0

Arithmetics

Powers

For all real numbers x, y and positive numbers a, b

- $a^x a^y = a^{x+y}$
- $\frac{a^x}{a^y} = a^{x-y}$
- $(ab)^x = a^x b^x$
- $(\frac{a}{b})^x = \frac{a^x}{b^x}$
- $\frac{1}{a^x} = a^{-x}$
- $(a^x)^y = a^{xy}$
- $a^0 = 1$
- $a^{\frac{1}{n}} = \sqrt[n]{a}$, where n is a positive integer

Natural logarithm

For positive numbers x, y

- $e^x = y \iff x = \ln(y)$
- $\ln(xy) = \ln(x) + \ln(y)$
- $\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$
- $\ln\left(x^p\right) = p\ln(x)$

Natural logarithm is also often written as $\log(x)$.

Some special functions

Factorial of positive integers n

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$$

and $0! = 1$.

Gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \mathrm{d}t$$

The Gamma function has the following properties

$\Gamma(n) = (n-1)!$ if n is a positive integer

$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for any $\alpha > 0$

Beta function is defined as

$$\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} \mathrm{d}t$$

The Gamma and Beta functions are related by

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Derivatives

Derivatives of elementary functions

k, n and a are constants.

- $\frac{\mathrm{d}}{\mathrm{d}x} k = 0$
- $\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}$
- $\frac{\mathrm{d}}{\mathrm{d}x} e^{ax} = ae^{ax}$
- $\frac{\mathrm{d}}{\mathrm{d}x} \ln(x) = \frac{1}{x},\ x > 0$
- $\frac{\mathrm{d}}{\mathrm{d}x} a^x = a^x \ln a$
- $\frac{\mathrm{d}}{\mathrm{d}x} \frac{1}{x} = -\frac{1}{x^2}$

Derivatives of combined functions

$f(x)$ and $g(x)$ are differentiable functions, and k a constant.

- Constant rule

$$\frac{\mathrm{d}}{\mathrm{d}x}(k \cdot f(x)) = k \cdot f'(x)$$

- Sum rule

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(x) + g(x)) = f'(x) + g'(x)$$

- Product rule

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(x) \cdot g(x)) = f'(x)g(x) + f(x)g'(x)$$

- Quotient rule

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{\left(g(x)\right)^2}$$

- Reciprocal rule

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{g(x)}\right) = -\frac{g'(x)}{\left(g(x)\right)^2}$$

- Chain rule for a composite function

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x)) \cdot g'(x)$$

Integrals

Anti-derivatives

C and k are constants.

- $\int x^n \,\mathrm{d}x = \frac{1}{n+1}x^{n+1} + C,\ n \neq -1$
- $\int e^{ax} \,\mathrm{d}x = \frac{1}{a}e^{ax} + C,\ a \neq 0$
- $\int \frac{1}{x} \,\mathrm{d}x = \ln|x|,\ x > 0$

Definite integral of $f(x)$ from a to b

$$\int_a^b f(x) \,\mathrm{d}x = [F(x)]_a^b = F(b) - F(a)$$

Integrals of combined functions

$f(x)$ and $g(x)$ are integrable functions and k a constant.

- Constant rule

$$\int k \cdot f(x) \,\mathrm{d}x = k \cdot \int f(x) \,\mathrm{d}x$$

- Sum rule

$$\int (f(x) + g(x)) \,\mathrm{d}x = \int f(x) \,\mathrm{d}x + \int g(x) \,\mathrm{d}x$$

Combinatorics

Combinations and Permutations

How many ways can we choose k elements from n elements?		
	with replacement	without replacement
order	n^k	${}_nP_k = \frac{n!}{(n-k)!}$
no order	not included	${}_nC_k = \binom{n}{k} = \frac{n!}{(n-k)!k!}$

Descriptive Statistics

Sample mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sample Variance

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Sample standard deviation

$$s_x = \sqrt{s_x^2}$$

Sample covariance

$$s_{xy} = \mathrm{Cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Sample correlation

$$r_{xy} = \mathrm{Corr}(x,y) = \frac{s_{xy}}{s_x s_y}$$

Basic Probability

Addition Rule

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

Multiplication Rule

$$P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{B}|\mathbf{A})P(\mathbf{A}) = P(\mathbf{A}|\mathbf{B})P(\mathbf{B})$$

Law of Total Probability - basic version

$$P(\mathbf{A}) = P(\mathbf{A}|\mathbf{B})P(\mathbf{B}) + P(\mathbf{A}|\mathbf{B}^c)P(\mathbf{B}^c)$$

where $\mathbf{B}^c$ is the complement of $\mathbf{B}$.

Law of Total Probability - general partition

$$P(\mathbf{A}) = \sum_{k=1}^{K} P(\mathbf{A}|\mathbf{B}_k)P(\mathbf{B}_k)$$

where $\mathbf{B}_1, \ldots, \mathbf{B}_K$ is a partitioning of the sample space.

Bayes' Theorem - basic version

$$P(\mathbf{B}|\mathbf{A}) = \frac{P(\mathbf{A}|\mathbf{B})P(\mathbf{B})}{P(\mathbf{A})}$$

Bayes' Theorem - general partition

$$P(\mathbf{B}_k|\mathbf{A}) = \frac{P(\mathbf{A}|\mathbf{B}_k)P(\mathbf{B}_k)}{P(\mathbf{A})}$$

Properties of One Random Variable

Expected value

If X is a continuous variable with density function $f(x)$

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\mathrm{d}x$$

Expected value of a function $g(X)$

If X is a continuous variable with density function $f(x)$

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\mathrm{d}x$$

Variance

If X is a continuous variable with density function $f(x)$

$$\sigma^2 = \mathbb{V}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\mathrm{d}x = \mathbb{E}(X^2) - \mu^2$$

Standard deviation

$$\sigma = \mathbb{S}(X) = \sqrt{\mathbb{V}(X)}$$

Expected value linear combination (c and d are constants)

$$\mathbb{E}(c + d \cdot X) = c + d \cdot \mathbb{E}(X)$$

Variance linear combination

$$\mathbb{V}(c + d \cdot X) = d^2 \cdot \mathbb{V}(X)$$

Distribution of a transformation

Let X be a continuous random variable and $Y = g(X)$, where $g(x)$ is a monotone differentiable function with inverse function $x = g^{-1}(y)$. Then,

$$f_Y(y) = f_X\Big(g^{-1}(y)\Big) \cdot \left|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right|$$

Properties of Two Random Variables

Expected value of a linear combination

$$\mathbb{E}(cX + dY) = c\mathbb{E}(X) + d\mathbb{E}(Y)$$

Variance for a linear combination

$$\mathbb{V}(cX + dY) = c^2\mathbb{V}(X) + d^2\mathbb{V}(Y) + 2cd\mathrm{Cov}(X,Y)$$

Marginal distribution for X

If X and Y are continuous variables with joint density function $f(x,y)$, then the marginal density of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y$$

Conditional distribution for Y given X

If X and Y are continuous variables with joint density function $f(x,y)$, then the conditional density of Y is

$$f(y \mid x) = \frac{f(x,y)}{f_X(x)}, \quad f_X(x) > 0$$

where $f_X(x)$ is the marginal density for X.

Law of iterated expectation

$$\mathbb{E}_Y(Y) = \mathbb{E}_X\Big(\mathbb{E}_{Y|X}(Y|X)\Big)$$

Law of total variance

$$\mathbb{V}_Y(Y) = \mathbb{E}_X\Big(\mathbb{V}_{Y|X}(Y|X)\Big) + \mathbb{V}_X\Big(\mathbb{E}_{Y|X}(Y|X)\Big)$$

Covariance between two random variables X and Y

$$\mathrm{Cov}(X,Y) = \mathbb{E}\Big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\Big) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Correlation between two random variables X and Y

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathbb{S}(X) \cdot \mathbb{S}(Y)}$$

Properties of the Sample Mean

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with expected value $\mu = E(X_i)$ and variance $\sigma^2 = \text{Var}(X_i)$. For the sample mean $\overline{X}_n = \sum_{i=1}^{n} X/n$ we have:

Expected value of the sample mean

$$\text{E}(\overline{X}_n) = \mu$$

Variance of the sample mean

$$\text{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$$

Law of Large Numbers

$$\overline{X}_n \xrightarrow{p} \mu$$

where $\xrightarrow{p}$ is convergence in probability: for all $\epsilon > 0$

$$P(|\overline{X}_n - \mu| > \epsilon) \to 0 \text{ when } n \to \infty$$

Central Limit Theorem (informally)

$$\overline{X}_n \overset{\text{approx}}{\sim} \text{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ for large } n$$

Rule of Thumb: the approximation is accurate if $n \geq 30$.

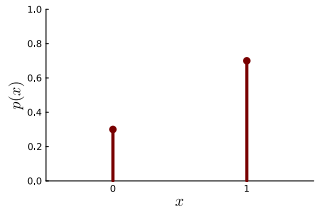
Discrete Distributions

Bernoulli distribution $X \sim \text{Bernoulli}(\theta)$

$$p(x) = \begin{cases} 1-\theta & \text{if } x = 0 \\ \theta & \text{if } x = 1 \end{cases}$$

$$\mathbb{E}(X) = \theta$$

$$\mathbb{V}(X) = \theta(1-\theta)$$

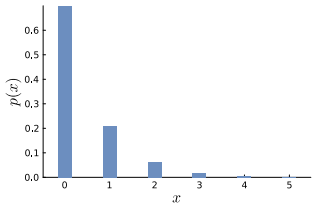


Geometric distribution $X \sim \text{Geom}(\theta)$

$$p(x) = (1-\theta)^x\theta \text{ for } x = 0, 1, 2, \ldots$$

$$\mathbb{E}(X) = \frac{1-\theta}{\theta}$$

$$\mathbb{V}(X) = \frac{1-\theta}{\theta^2}$$

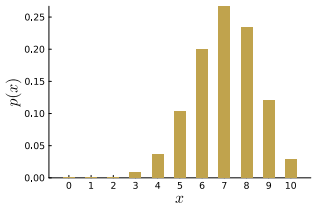


Binomial Distribution: $X \sim \text{Binomial}(n, \theta)$

$$p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

$$\mathbb{E}(X) = n\theta$$

$$\mathbb{V}(X) = n\theta(1-\theta)$$

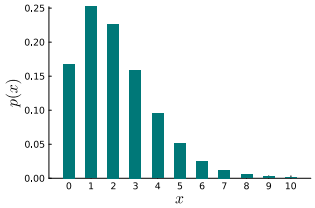


Negative Binomial distribution: $X \sim \text{NegBin}(r, \theta)$

$$p(x) = \binom{x+r-1}{x}\theta^r(1-\theta)^x \text{ for } x = 0, 1, 2, \ldots$$

$$\mathbb{E}(X) = \frac{r(1-\theta)}{\theta}$$

$$\mathbb{V}(X) = \frac{r(1-\theta)}{\theta^2}$$

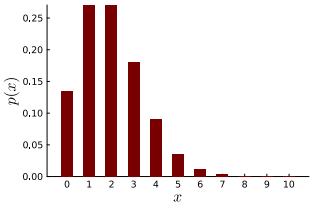


Poisson Distribution: $X \sim \text{Pois}(\theta)$

$$p(x) = \frac{e^{-\theta}\theta^x}{x!} \text{ for } x = 0, 1, 2, \ldots$$

$$\mathbb{E}(X) = \theta$$

$$\mathbb{V}(X) = \theta$$



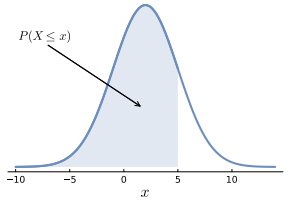
Continuous Distributions

Normal Distribution: $X \sim \text{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \text{ for } -\infty < x < \infty$$

$$\mathbb{E}(X) = \mu$$

$$\mathbb{V}(X) = \sigma^2$$



If $X \sim \text{N}(\mu, \sigma^2)$ and $Y = c + d \cdot X$ then

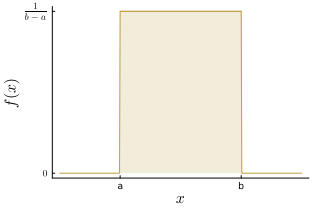
$$Y \sim N\left(c + d \cdot \mu,\ d^2 \cdot \sigma^2\right)$$

Uniform distribution: $X \sim \text{U}(a, b)$

$$f(x) = \frac{1}{b-a}, \text{ for } a \leq x \leq b$$

$$\mathbb{E}(X) = (a+b)/2$$

$$\mathbb{V}(X) = (b-a)^2/12$$

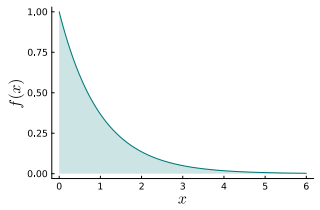


Exponential distribution: $X \sim \text{Expon}(\theta)$

$$f(x) = \theta\, e^{-\theta x}, \text{ for } x \geq 0$$

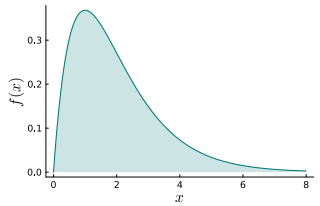
$$\mathbb{E}(X) = \frac{1}{\theta}$$

$$\mathbb{V}(X) = \frac{1}{\theta^2}$$



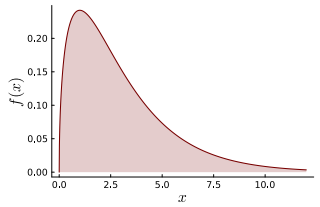
Gamma distribution $X \sim \text{Gamma}(\alpha, \beta)$

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \text{ for } x \geq 0$$
$$\mathbb{E}(X) = \frac{\alpha}{\beta}$$
$$\mathbb{V}(X) = \frac{\alpha}{\beta^2}$$



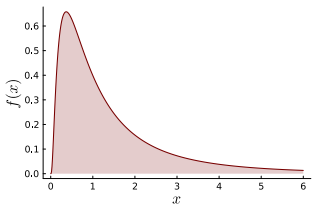
Inv-χ^2 distribution $X \sim \text{Inv-}\chi^2(\nu, \tau^2)$

$$f(x) = \frac{(\tau^2\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{\exp\left(\frac{-\nu\tau^2}{2x}\right)}{x^{1+\nu/2}}, \text{ for } x \geq 0$$
$$\mathbb{E}(X) = \frac{\nu}{\nu-2}\tau^2$$
$$\mathbb{V}(X) = \frac{2\nu^2\tau^4}{(\nu-2)^2(\nu-4)}$$



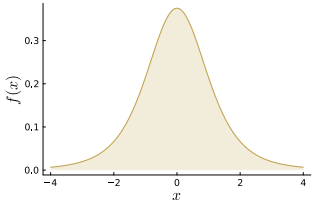
Log-normal distribution $X \sim \text{LogNormal}(\mu, \sigma^2)$

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2} \text{ for } 0 < x < \infty$$
$$\mathbb{E}(X) = \exp(\mu + \sigma^2/2)$$
$$\mathbb{V}(X) = \Big(\exp(\sigma^2) - 1\Big)\exp(2\mu + \sigma^2)$$



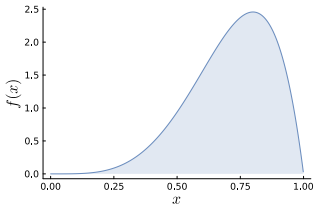
Student t-distribution $X \sim t(\nu)$

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, -\infty < x < \infty$$
$$\mathbb{E}(X) = 0 \text{ if } \nu > 1$$
$$\mathbb{V}(X) = \frac{\nu}{\nu-2} \text{ if } \nu > 2$$



Beta distribution $X \sim \text{Beta}(\alpha, \beta)$

$$f(x) = \frac{1}{\text{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \text{ for } 0 < x < 1$$
$$\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$$
$$\mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
where
$$\text{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$



Multivariate Distributions

Multivariate normal distribution

$$(Y_1, Y_2, \ldots, Y_p)^\top \sim \text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

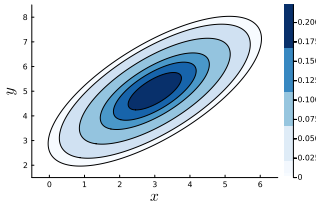
where $\boldsymbol{\mu}$ is the p-element mean vector and $\boldsymbol{\Sigma}$ is the $p \times p$ covariance matrix.

In the bivariate case with $p = 2$:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

and ρ_{12} is the correlation between Y_1 and Y_2.



Multinomial Distribution:

$$(X_1, \ldots, X_p)^\top \sim \text{Multinom}(n, \theta_1, \ldots, \theta_p)$$

$$f(x_1, \ldots, x_p) = \frac{n!}{x_1! \cdots x_p!}\theta_1^{x_1} \cdots \theta_p^{x_p}$$
for $x_k \geq 0, \sum_{k=1}^p x_k = n$
$$\mathbb{E}(X_k) = n\theta_k$$
$$\mathbb{V}(X_k) = n\theta_k(1-\theta_k)$$

Dirichlet distribution

$(X_1, \dots, X_p)^\top \sim \text{Dirichlet}(\alpha_1, \alpha_2, \dots, \alpha_p)$

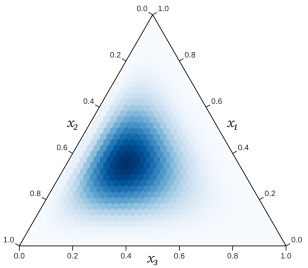
$f(x_1, \dots, x_p) \propto x_1^{\alpha_1 - 1} \cdots x_p^{\alpha_p - 1}$,

for $0 < x_k < 1$ and $\sum_{k=1}^p x_k = 1$

$\mathbb{E}(X_j) = \frac{\alpha_j}{\alpha_+}$

$\mathbb{V}(X_j) = \frac{\mathbb{E}(X_j)(1 - \mathbb{E}(X_j))}{1 + \alpha_+}$

where $\alpha_+ = \sum_{k=1}^p \alpha_k$



Likelihood and Information

Log-likelihood $\ell(\theta)$

If $Y_1, Y_2, \dots, Y_n$ are *iid* with density/probability function $p(y_i \,|\, \theta)$

$$\ell(\theta) = \sum_{i=1}^n \log p(y_i \,|\, \theta)$$

Maximum likelihood estimator (MLE) $\hat{\theta}$

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ell(\theta)$$

Observed information

$$J_{\mathbf{y}}(\hat{\theta}) = -\ell''(\hat{\theta})$$

Fisher information

$$I(\theta) = \mathbb{E}_{\mathbf{Y}|\theta}(-\ell''(\theta))$$

Observed posterior information

$$J_{\mathbf{y}}(\tilde{\theta}) = -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log p(\theta \,|\, \mathbf{y})\Big|_{\theta = \tilde{\theta}}$$

where $\tilde{\theta}$ is the posterior mode

$$\tilde{\theta} = \arg\max_{\theta \in \Theta} p(\theta \,|\, \mathbf{y})$$

Jeffreys' prior

$$p(\boldsymbol{\theta}) \propto |I(\theta)|^{1/2}$$

Normal posterior approximation (informally)

$$\theta | \mathbf{y} \overset{\text{approx}}{\sim} N\left(\tilde{\theta}, J_{\mathbf{y}}^{-1}(\tilde{\theta})\right) \text{ for large } n$$

IID Gaussian - variance known

Model

$$X_1, \dots, X_n \,|\, \theta, \sigma^2 \overset{\text{iid}}{\sim} N(\theta, \sigma^2), \quad \sigma^2 \text{ known.}$$

Conjugate prior

$$\theta \sim N(\mu_0, \tau_0^2)$$

Posterior

$$\theta \,|\, x_1, \dots, x_n \sim N(\mu_n, \tau_n^2)$$

$\mu_n = w\bar{x} + (1 - w)\mu_0$

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

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

$\bar{x} = n^{-1} \sum_{i=1}^n x_i$

Predictive distribution

$$\tilde{x} \,|\, x_1, \dots, x_n \sim N(\mu_n, \sigma^2 + \tau_n^2)$$

IID Gaussian - both parameters unknown

Model

$$X_1, \dots, X_n \,|\, \theta, \sigma^2 \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$$

Conjugate prior

$$\theta \,|\, \sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$$
$$\sigma^2 \sim \text{Inv}{-}\chi^2(\nu_0, \sigma_0^2).$$

Joint posterior

$$\theta \,|\, \sigma^2, x_1, \dots, x_n \sim N(\mu_n, \sigma^2/\kappa_n)$$
$$\sigma^2 \,|\, x_1, \dots, x_n \sim \text{Inv}{-}\chi^2(\nu_n, \sigma_n^2)$$

$\mu_n = w\bar{x} + (1 - w)\mu_0$

$w = \frac{n}{\kappa_0 + n}$

$\kappa_n = \kappa_0 + n$

$\nu_n = \nu_0 + n$

$\nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{x} - \mu_0)^2$

$\bar{x} = n^{-1} \sum_{i=1}^n x_i$

$(n-1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$

Marginal posterior for θ

$$\theta \,|\, x_1, \dots, x_n \sim t\left(\boldsymbol{\mu}_n, \frac{\sigma_n^2}{\kappa_n}, \nu_n\right)$$

Linear Gaussian regression

Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \varepsilon \overset{iid}{\sim} N(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

Conjugate prior

$$\begin{aligned}\boldsymbol{\beta}|\sigma^2 &\sim N(\boldsymbol{\mu}_0, \sigma^2\boldsymbol{\Omega}_0^{-1}) \\ \sigma^2 &\sim \text{Inv}{-}\chi^2(\nu_0, \sigma_0^2)\end{aligned}$$

Joint posterior

$$\begin{aligned}\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X} &\sim N(\boldsymbol{\mu}_n, \sigma^2\boldsymbol{\Omega}_n^{-1}) \\ \sigma^2|\mathbf{y}, \mathbf{X} &\sim \text{Inv}{-}\chi^2(\nu_n, \sigma_n^2)\end{aligned}$$

$\boldsymbol{\Omega}_n = \mathbf{X}^\top\mathbf{X} + \boldsymbol{\Omega}_0$
$\boldsymbol{\mu}_n = \boldsymbol{\Omega}_n^{-1}(\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Omega}_0\boldsymbol{\mu}_0)$
$\nu_n = \nu_0 + n$
$\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + (n-p)s^2 + (\boldsymbol{\mu}_n - \boldsymbol{\mu}_0)^\top\boldsymbol{\Omega}_0(\boldsymbol{\mu}_n - \boldsymbol{\mu}_0) + (\boldsymbol{\mu}_n - \hat{\boldsymbol{\beta}})^\top\mathbf{X}^\top\mathbf{X}(\boldsymbol{\mu}_n - \hat{\boldsymbol{\beta}})$
$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$
$s^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-p)$

Marginal posterior for $\boldsymbol{\beta}$

$$\boldsymbol{\beta}|\mathbf{y} \sim t(\boldsymbol{\mu}_n, \sigma_n^2\boldsymbol{\Omega}_n^{-1}, \nu_n)$$

Marginal posterior individual coefficients

$$\beta_j|\mathbf{y} \sim t\left(\mu_{n,j}, \sigma_{\beta,j}^2, \nu_n\right)$$

$\mu_{n,j}$ is the jth element of $\boldsymbol{\mu}_n$ and
$\sigma_{\beta,j}^2$ is the jth diagonal element of $\sigma_n^2\boldsymbol{\Omega}_n^{-1}$.

Model comparison

Laplace approximation of the marginal likelihood

$$\log\hat{p}(\mathbf{y}) = \log p(\mathbf{y}|\tilde{\boldsymbol{\theta}}) + \log p(\tilde{\boldsymbol{\theta}}) + \frac{1}{2}\log\left|J_{\mathbf{y}}^{-1}(\tilde{\boldsymbol{\theta}})\right| + \frac{p}{2}\ln(2\pi)$$

where p is the number of parameters in the model.

Log predictive density score (LPDS)

$$\text{LPDS} = \log p(\mathbf{y}_{\text{test}}|\mathbf{y}_{\text{train}})$$

where

$$p(\mathbf{y}_{\text{test}}|\mathbf{y}_{\text{train}}) = \int p(\mathbf{y}_{\text{test}}|\mathbf{y}_{\text{train}}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{\text{train}})d\boldsymbol{\theta}$$

Leave-one-out LPDS

$$\text{LPDS}_{\text{LOO}} = \sum_{i=1}^{n}\log p(y_i|\mathbf{y}_{-i})$$

where

$$p(y_i|\mathbf{y}_{-i}) = \int p(y_i|\mathbf{y}_{-i}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{-i})d\boldsymbol{\theta}$$

and $\mathbf{y}_{-i}$ is the full dataset, except the ith observation.

Posterior sampling algorithms

Random Walk Metropolis algorithm

Initial value $\boldsymbol{\theta}^{(0)}$
for i in $1{:}(m{+}b)$ **do**
  Draw proposal
  $\boldsymbol{\theta}^\star \sim N(\boldsymbol{\theta}^\star|\boldsymbol{\theta}^{(i-1)}, c\cdot\Sigma)$
  $\alpha \leftarrow \min\left(1, \frac{p(\mathbf{y}|\boldsymbol{\theta}^\star)p(\boldsymbol{\theta}^\star)}{p(\mathbf{y}|\boldsymbol{\theta}^{(i-1)})p(\boldsymbol{\theta}^{(i-1)})}\right)$
  Draw $u \sim \text{Uniform}(0,1)$
  if $u < \alpha$ **then**
    $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^\star$
  else
    $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$
  end
end

Created by Mattias Villani, Stockholm University.