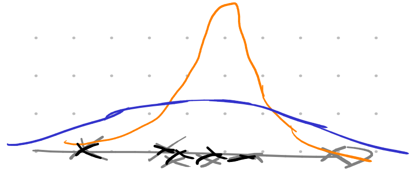
Байесовский вывод для гауссиана

$$y \sim \mathcal{N}(\bar{w}^T \bar{x}, \sigma^2) \quad \sigma^2 = \text{Const}$$



$$D = \{x_1 \ldots x_n\}$$

$$p(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(x-\mu)^2}$$

"τ precision"

$\Lambda = \Sigma^{-1}$ precision matrix

$$\mathcal{N}(x_1 \ldots x_n \mid \mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} e^{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\log p(D \mid \mu, \tau) = -\frac{n}{2}\log 2\pi + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

① $\tau = \text{Const}$, $\mu = ?$

$$\log p(D \mid \mu) = \text{Const} - \frac{\tau}{2}\sum_i (x_i-\mu)^2$$

$$p(\mu) = \mathcal{N}(\mu \mid \mu_0, \tau_0) = \sqrt{\frac{\tau_0}{2\pi}}\, e^{-\tau_0/2\,(\mu-\mu_0)^2}$$

$$\log p(\mu \mid D) = \text{Const} - \frac{\tau}{2}\sum_i (x_i-\mu)^2 - \frac{\tau_0}{2}(\mu-\mu_0)^2 =$$

$$= \text{Const} - \frac{\mu^2}{2}(\tau_0 + n\cdot\tau) + \mu\cdot(\tau_0\mu_0 + \tau\cdot\sum x_i) =$$

$$= \text{Const} - \frac{\tau_0 + n\tau}{2}\left(\mu - \frac{\tau_0\mu_0 + \tau\sum x_i}{\tau_0 + n\tau}\right)^2$$

$$\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$$

$$p(\mu \mid x_1 \ldots x_n) = \mathcal{N}\left(\mu \,\middle|\, \frac{\tau_0\mu_0 + \tau\sum_i x_i}{\tau_0 + n\tau}, \tau_0 + \tau n\right)$$

② $\mu = \text{Const}$, $\tau = ?$

$$p \propto \tau^{\ldots} \cdot e^{\tau \ldots}$$

$$\log p(D \mid \tau) = \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_i (x_i-\mu)^2$$

$$p(\tau \mid \alpha_0, \beta_0) = \boxed{\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}}\, \tau^{\alpha_0 - 1} e^{-\beta_0\tau}$$

— gamma distribution

$\tau \in (0, +\infty)$

$$\log p(\tau \mid D) = \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_i (x_i-\mu)^2$$

$$+ (\alpha_0 - 1)\log\tau - \beta_0\tau = \text{Const} + \left(\alpha_0 + \frac{n}{2} - 1\right)\log\tau - \tau\left(\beta_0 + \sum_i (x_i-\mu)^2\right)$$

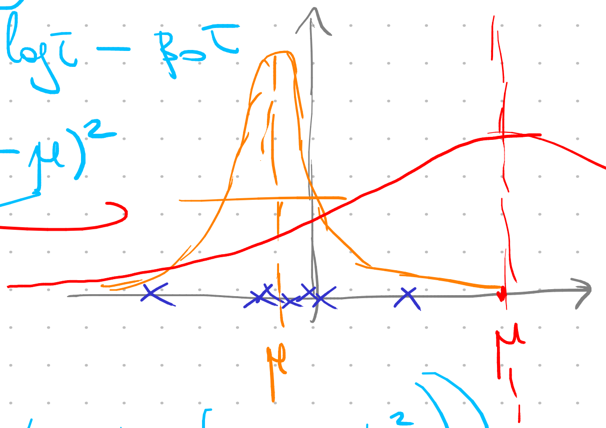
$$p(\tau \mid x_1 \ldots x_n) = \text{Gamma}\left(\tau \,\middle|\, \alpha_0 + \frac{n}{2}, \beta_0 + \sum_i (x_i-\mu)^2\right)$$

③ $\mu, \tau = ?$ $\qquad p(\mu,\tau) = ?$

$$p(\mu,\tau) = p(\mu)p(\tau) = \mathcal{N}(\mu|\mu_0,\tau_0)\,\mathrm{Gam}(\tau|\alpha_0,\beta_0)$$

$$\log p(\mu,\tau|D) = \text{Const} - \frac{\tau_0}{2}(\mu-\mu_0)^2 + (\alpha_0-1)\log\tau - \beta_0\tau$$

$$+ \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$



$\mathrm{Gam}(\tau|\alpha_0,\beta_0)$

$$p(\mu,\tau) = p(\tau)\,p(\mu|\tau) = p(\mu)\,p(\tau|\mu)$$

$$\mathcal{N}(\mu|\mu_0, \lambda_0\tau) \qquad \mathcal{N}(\mu|\mu_0,\tau_0)\,\mathrm{Gam}(\tau|\alpha_0, \beta_0\cdot(\ldots\ldots\mu^2))$$

$\lambda\tau$

$$\log p(\mu,\tau) = \text{Const} + (\alpha_0-1)\log\tau - \beta_0\tau + \boxed{\frac{1}{2}\log\tau} - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2$$

$$+ \log p(D|\mu,\tau)$$

$$\log p(\mu,\tau|D) = \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_i(x_i-\mu)^2 + \ldots =$$

$$= \text{Const} + (\alpha_0 + \frac{n}{2} - 1)\log\tau + \frac{1}{2}\log\tau - \beta_0\tau - \frac{\tau}{2}\left(\lambda_0(\mu-\mu_0)^2 + \sum_i(x_i-\mu)^2\right)$$

$$\lambda_0\mu^2 - 2\lambda_0\mu\mu_0 + \lambda_0\mu_0^2 + \sum_i x_i^2 - 2\mu\sum x_i + n\cdot\mu^2$$

$$= (\lambda_0+n)\mu^2 - 2\mu(\lambda_0\mu_0 + \sum x_i) + \lambda_0\mu_0^2 + \sum_i x_i^2$$

$$= \underbrace{(\lambda_0+n)\left(\mu - \frac{\lambda_0\mu_0+\sum x_i}{\lambda_0+n}\right)^2}_{\text{Gaussian}} + \underbrace{\lambda_0\mu_0^2 + \sum_i x_i^2 - \frac{(\lambda_0\mu_0+\sum x_i)^2}{\lambda_0+n}}_{\beta\ /\ \text{Gamma}}$$

$$= \text{Const} + (\alpha_0+\frac{n}{2}-1)\log\tau + \frac{1}{2}\log\tau - \frac{\tau(\lambda_0+n)}{2}\left(\mu - \frac{\lambda_0\mu_0+\sum x_i}{\lambda_0+n}\right)^2$$

$$-\tau(\beta_0 + \ldots)$$

$\tau \sim \mathrm{Gam}(\_\_)$

$\sigma^2 \sim$ Inverse Gamma

$$p(\tau) = \text{Gam}(\tau \mid \alpha_0, \beta_0)$$

$$p(\mu \mid \tau) = \mathcal{N}(\mu \mid \mu_0, \tau \cdot \lambda_0)$$

Conjugate priors

$$\times \quad p(x_1 \ldots x_n \mid \mu, \tau) \quad \propto$$

$$E[p(\tau)] = \alpha/\beta, \quad \text{Mode} = \frac{\alpha-1}{\beta}$$

$$\propto \quad p(\mu, \tau \mid D) = \text{Gam}(\tau \mid \alpha_n, \beta_n) \cdot \mathcal{N}(\mu \mid \mu_n, \lambda_n \tau), \text{ where}$$

$$\lambda_n = \lambda_0 + n \qquad \alpha_n = \alpha_0 + n/2$$

$$\mu_n = \frac{\lambda_0 \mu_0 + \sum_i x_i}{\lambda_0 + n} \qquad \beta_n = \beta_0 + \lambda_0 \mu_0^2 + \sum x_i^2 - \frac{(\lambda_0 \mu_0 + \sum_i x_i)^2}{\lambda_0 + n}$$

(4) Posterior for $\mu = ?$ — Marginalization w.r.t. τ

$$p(\mu \mid x_1 \ldots x_n) = \int_0^\infty p(\mu, \tau \mid D)\, d\tau =$$

$$\frac{\beta^\alpha}{\Gamma(\alpha)}$$

$$= \int_0^\infty \text{Gam}(\tau \mid \alpha_n, \beta_n)\, \mathcal{N}(\mu \mid \mu_n, \lambda_n \tau)\, d\tau =$$

$$= \text{Const} \cdot \int_0^\infty \tau^{\alpha_n - 1} e^{-\beta_n \tau} \cdot \tau^{\frac{1}{2}} \cdot e^{-\frac{\lambda_n \tau}{2}(\mu - \mu_n)^2}\, d\tau =$$

$$= \text{Const} \cdot \underbrace{\int_0^\infty \tau^{(\alpha_n + \frac{1}{2}) - 1} \cdot e^{-\tau \cdot \left(\beta_n + \frac{\lambda_n}{2}(\mu - \mu_n)^2\right)}\, d\tau}_{= \Gamma(\ldots)} =$$

$$= \text{Const} \cdot \frac{\Gamma(\alpha_n + 1/2)}{\left(\beta_n + \frac{\lambda_n}{2}(\mu - \mu_n)^2\right)^{\alpha_n + \frac{1}{2}}} = p(\mu \mid x_1 \ldots x_n)$$

fat heavy tails

Student's t distribution

Gosset (1890-s 1900-s)

(1870s)

$$p(t \mid D) = \frac{\Gamma\left(\frac{D+1}{2}\right)}{\sqrt{\pi D}\, \Gamma(D/2)} \left(1 + \frac{t^2}{D}\right)^{-\frac{D+1}{2}}$$

degrees of freedom

$$p(\mu | x_1 \dots x_n) = \frac{\text{Const}}{\left(1 + \frac{\lambda_n}{2\beta_n}(\mu - \mu_n)^2\right)^{\alpha_n + \frac{1}{2}}}$$

$$= \frac{\text{Const}}{\left(1 + \frac{1}{\alpha_n}\left(\sqrt{\frac{\alpha_n \lambda_n}{2\beta_n}}\,\mu - \mu_n \dots\right)^2\right)^{\alpha_n + 1/2}}$$

5 Predictive distribution

$$p(x | x_1 \dots x_n) = \iint p(x|\mu,\tau)\, p(\mu,\tau | x_1 \dots x_n)\, d\mu\, d\tau =$$

$$= \int_0^\infty \int_{-\infty}^\infty \mathcal{N}(x | \mu, \tau) \cdot \text{Gam}(\tau | \alpha_n, \beta_n) \cdot \mathcal{N}(\mu | \mu_n, \lambda_n \tau)\, d\mu\, d\tau =$$

$$\int_{-\infty}^\infty \sqrt{\frac{\tau}{2\pi}} \cdot e^{-\frac{\tau}{2}(x-\mu)^2} \sqrt{\frac{\tau_n}{2\pi}}\, e^{-\frac{\lambda_n \tau}{2}(\mu - \mu_n)^2}\, d\mu =$$

$$= \text{Const} \cdot \sqrt{\tau} \cdot \int_{-\infty}^\infty e^{-\frac{\tau}{2}\left(\mu^2(\lambda_n + 1) - 2\mu(x + \lambda_n \mu_n) + x^2 + \mu_n^2 \lambda_n\right)} d\mu$$

$$(\lambda_n + 1)\left(\mu - \frac{x + \lambda_n \mu_n}{\lambda_n + 1}\right)^2 -$$

$$= \text{Const} \cdot \cancel{\sqrt{\tau}} \cdot e^{-\frac{\tau}{2}\left(x^2 + \mu_n^2 \lambda_n - \frac{(\sim)^2}{\lambda_n + 1}\right)} \cdot \underbrace{\int e^{-\frac{\tau(\lambda_n + 1)}{2}\left(\mu - \frac{-}{-}\right)^2} d\mu}_{= \frac{\text{Const}}{\cancel{\sqrt{\tau}}}}$$

$$= \text{Const} \cdot e^{-\tau/2\left(\quad\right)} = (*)$$

$1 - \frac{1}{\dots}$

$$x^2 + \lambda_n \mu_n^2 - \frac{1}{\lambda_n + 1}\left(x^2 + 2x \lambda_n \mu_n + \lambda_n^2 \mu_n^2\right) =$$

$$= \frac{\lambda_n}{\lambda_n + 1} \cdot \mu_n^2$$

$$= + \frac{\lambda_n}{\lambda_n + 1} \cdot x^2 - \frac{2\lambda_n \mu_n}{\lambda_n + 1} x + \lambda_n \mu_n^2 - \frac{\lambda_n}{\lambda_n + 1} \cdot \lambda_n \mu_n^2 =$$

$$= + \frac{\lambda_n}{\lambda_n+1}(x - \mu_n)^2$$

$$\circledast = \text{Const} \cdot e^{-\frac{\lambda_n}{\lambda_n+1} \cdot \frac{\tau}{2}(x-\mu_n)^2}$$

$$= \int_0^\infty \mathrm{Gam}(\tau \mid \alpha_n, \beta_n) \cdot \mathcal{N}\left(x \mid \mu_n, \frac{\lambda_n}{\lambda_n+1}\tau\right) d\tau$$

$$p(x \mid D) = \frac{\text{Const}}{\left(1 + \frac{\lambda_n}{2(\lambda_n+1)\beta_n}(\mu - \mu_n)^2\right)^{\alpha_n + 1/2}}$$