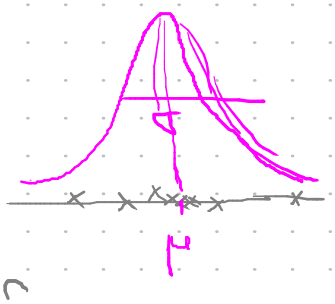
① Gaussians

$D = \{x_1, x_2, \dots, x_N\}$



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

$$\frac{1}{\sigma^2} = \tau \quad \text{precision}$$

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

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

$$p(x_1 \dots x_N \mid \mu, \tau) = \prod_n \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(x_n-\mu)^2}$$

$$\log p(x_1 \dots x_N \mid \mu, \tau) = -\frac{N}{2}\log 2\pi + \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2$$

② τ = Const, μ = ?

$$\log p(x_1 \dots x_N \mid \mu) = \text{Const} - \frac{\tau}{2}\sum_n (x_n-\mu)^2$$

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

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

$$= \text{Const} - \mu^2\left(\frac{\tau}{2}\cdot N + \frac{\tau_0}{2}\right) + \mu\left(\tau \cdot \Sigma x_n + \tau_0\mu_0\right) =$$

$$= \text{Const} - \frac{\tau_0 + N\tau}{2}\left(\mu - \frac{\tau_0\mu_0 + \tau\cdot\Sigma x_n}{\tau_0 + N\tau}\right)^2$$

$$p(\mu \mid x_1 \dots x_N) = \mathcal{N}\left(\mu \mid \frac{\tau_0\mu_0 + \tau\cdot\Sigma x_n}{\tau_0 + N\tau},\ \tau_0 + N\tau\right)$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

(3) $\mu = \text{const}$, $\tau = ?$

$$\log p(D|\tau) = \text{const} + \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_n (x_n - \mu)^2$$

$$p(\tau \mid \alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \tau^{\alpha_0 - 1} e^{-\beta_0 \tau}, \quad \tau \in (0, \infty)$$

– gamma distribution

$$\log p(\tau|D) = \text{const} + \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_n (x_n - \mu)^2 + (\alpha_0 - 1)\log\tau - \beta_0\tau$$

$$p(\tau|D) = \text{Gam}\left(\tau \mid \alpha_0 + \frac{N}{2}, \beta_0 + \frac{1}{2}\sum_n (x_n - \mu)^2\right)$$

(4) $\mu, \tau = ?$

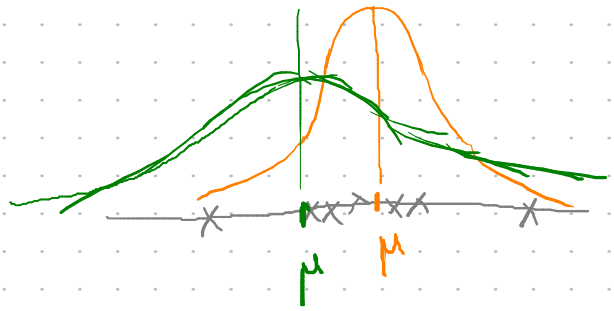
$$\log p(D|\mu,\tau) = \text{const} - \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_n (x_n - \mu)^2 = \lambda_0\tau$$

$$\text{const} - \frac{\tau\lambda_0}{2}(\mu - \mu_0)^2$$

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

$$p(\mu,\tau) = \mathcal{N}(\mu \mid \mu_0, \tau\lambda_0) \cdot \text{Gam}(\tau \mid \alpha_0, \beta_0)$$

$$\log p(\mu,\tau) = (\ldots\mu\ldots) + (\ldots\tau\ldots)$$



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

$$= \sqrt{\frac{\tau\lambda_0}{2\pi}}\, e^{-\ldots}$$

$$p(\mu,\tau) = \text{Gam}(\tau \mid \alpha_0, \beta_0) \cdot \mathcal{N}(\mu \mid \mu_0, \lambda_0\tau)$$

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

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

$$= \text{const} + (\alpha_0 - 1)\log\tau - \beta_0\tau - \frac{\lambda_0\tau}{2}(\mu - \mu_0)^2 + \frac{1}{2}\log\tau + \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_n (x_n - \mu)^2$$

$$p(\mu,\tau|D) = \text{Gam}(\tau \mid \alpha_N, \beta_N) \cdot \mathcal{N}(\mu \mid \mu_N, \lambda_N\tau)$$

$$-\frac{\tau}{2}\left(\lambda_0(\mu-\mu_0)^2+\sum_n(x_n-\mu)^2\right)=$$

$$=-\frac{\tau}{2}\left(\lambda_0\mu^2-2\lambda_0\mu_0\cdot\mu+\lambda_0\mu_0^2+N\cdot\mu^2-2\mu\cdot\sum x_n+\sum x_n^2\right)$$

$$=-\frac{\tau}{2}\left((\lambda_0+N)\left(\mu-\frac{\lambda_0\mu_0+\sum_n x_n}{\lambda_0+N}\right)^2-\frac{(\lambda_0\mu_0+\sum_n x_n)^2}{\lambda_0+N}+\lambda_0\mu_0^2+\sum_n x_n^2\right)$$

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

$$-\frac{\tau}{2}\left(\quad\right)=$$

$\log \mathcal{N}(\mu|\mu_N,\lambda_N\cdot\tau)$

$$=\text{Const}+\frac{1}{2}\log\tau-\frac{\tau}{2}(\lambda_0+N)\left(\mu-\frac{\lambda_0\mu_0+\sum_n x_n}{\lambda_0+N}\right)^2$$

$$+(\alpha_N-1)\log\tau-\beta_N\cdot\tau$$

$$\alpha_N=\alpha_0+\frac{N}{2};\quad \beta_N=\beta_0+\frac{1}{2}\left(\lambda_0\mu_0^2+\sum x_n^2-\frac{(\lambda_0\mu_0+\sum_n x_n)^2}{\lambda_0+N}\right)$$

$$\mu_N=\frac{\lambda_0\mu_0+\sum_n x_n}{\lambda_0+N};\quad \lambda_N=\lambda_0+N$$

$p(\sigma^2)=$ Inverse Gamma

5 Posterior for $\mu=?$

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

$$=\int_0^\infty \text{Gam}(\tau|\alpha_N,\beta_N)\cdot\mathcal{N}(\mu|\mu_N,\lambda_N\cdot\tau)\,d\tau=$$

$$= \text{const} \cdot \int_0^\infty \tau^{\alpha_N - 1} e^{-\beta_N \tau} \cdot \tau^{\frac{1}{2}} \cdot e^{-\frac{\lambda_N}{2}\tau(\mu-\mu_N)^2} d\tau$$

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

$$= \text{const} \cdot \int_0^\infty \tau^{(\alpha_N + \frac{1}{2}) - 1} \cdot e^{-\left(\beta_N + \frac{\lambda_N}{2}(\mu-\mu_N)^2\right)} d\tau =$$

$$= \text{const} \cdot \frac{\Gamma\left(\alpha_N + \frac{1}{2}\right)}{\left(\beta_N + \frac{\lambda_N}{2}(\mu-\mu_N)^2\right)^{\alpha_N + \frac{1}{2}}}$$

const

$-\frac{2\alpha_N + 1}{2}$

fat tails

Student's t-distribution

degrees of freedom

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

Gosset / Student - 1890s / 1900s

1870s —

(6) Predictive distribution

$$p(x|x_1, x_2, \dots, x_N) = \iint p(x|\mu,\tau)\, p(\mu,\tau|x_1, x_2, \dots, x_N)\, d\mu\, d\tau$$

$$= \iint \mathcal{N}(x|\mu,\tau) \cdot \mathcal{N}(\mu|\mu_N, \lambda_N\tau) \cdot \text{Gam}(\tau|\alpha_N, \beta_N)\, d\mu\, d\tau$$

$$\int_{-\infty}^{\infty} \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(x-\mu)^2} \cdot \sqrt{\frac{\lambda_N\tau}{2\pi}} e^{-\frac{\lambda_N\tau}{2}(\mu-\mu_N)^2} d\mu =$$

$$\left(= f(x) \cdot \int \mathcal{N}(\mu|\mu', \tau)\, d\mu \right)$$

$$= \text{Const} \cdot \sqrt{\tau} \left(\int \sqrt{\frac{\tau \cdot \lambda'}{2\pi}} \cdot e^{-\frac{\tau \cdot \lambda'}{2}(\mu - \mu')^2} \, d\mu \right) \cdot e^{-\frac{\tau}{2}(\dots x^2 \dots)}$$

$$-\frac{\tau}{2}\left(x^2 - 2\mu x + \mu^2 - \lambda_N \left(\mu^2 - 2\mu\mu_N + \mu_N^2 \right) \right) =$$

$$= -\frac{\tau(\lambda_N + 1)}{2}\left(\mu - \frac{\lambda_N \mu_N + x}{\lambda_N + 1} \right)^2 - \frac{\tau}{2}\left(x^2 + \mu_N^2 \lambda_N - \frac{(\lambda_N \mu_N + x)^2}{\lambda_N + 1} \right)$$

$$= \int \text{Gam}(\tau \mid \alpha_N, \beta_N) \cdot N\left(x \mid \mu_N, \frac{\lambda_N}{\lambda_N + 1} \cdot \tau \right) d\tau = (*)$$

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

$$\lambda_N \left(1 - \frac{\lambda_N}{\lambda_N + 1} \right) = \frac{\lambda_N}{\lambda_N + 1}$$

$$= \frac{\lambda_N}{\lambda_N + 1} x^2 - 2x \frac{\lambda_N}{\lambda_N + 1} \mu_N + \lambda_N \mu_N^2 - \frac{\lambda_N}{\lambda_N + 1} \cdot \lambda_N \cdot \mu_N^2 =$$

$$= \frac{\lambda_N}{\lambda_N + 1}\left(x^2 - 2x\mu_N + \cdot \mu_N^2 \right) = \frac{\lambda_N}{\lambda_N + 1} \cdot (x - \mu_N)^2$$

$$(*) = \frac{\text{Const}}{\left(\beta_N + \frac{\lambda_N}{2(\lambda_N + 1)} (x - \mu_N)^2 \right)^{\alpha_N + \frac{1}{2}}}$$