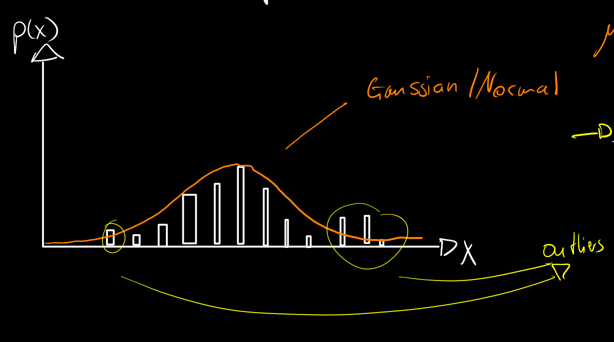
Posterior & MAP for (univariate) Normal with unknown mean & precision



μ & $\sigma|\sigma^2|\tau$ unknown

→ MLE ?

→ robuster estimate?

→ prior knowledge

we want regularize both of the paramaters

μ $\quad$ σ, σ^2, τ

↳ we need a prior over μ & τ

Likelihood

$$\mathcal{L}(D;\mu,\tau) = \prod_{i=0}^{N-1} \mathcal{N}(x^{[i]} | \mu, \tau^{-1})$$

$$= \prod_{i=0}^{N-1} \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2} \cdot (x^{[i]} - \mu)^2\right)$$

$$= \frac{\tau^{N/2}}{(2\pi)^{N/2}} \exp\left(-\frac{\tau}{2} \underbrace{\sum_{i=0}^{N-1} (x^{[i]} - \mu)^2}_{\text{sum of squared differences from the mean}}\right)$$

$$\sim \tau^{N/2} \exp\left(-\frac{\tau}{2} \cdot \left(\sum_{i=0}^{N-1} (x^{[i]} - \bar{x}) + N(\bar{x} - \mu)^2\right)\right)$$

$$\bar{x} = \frac{1}{N} \sum_{i=0}^{N-1} x^{[i]} = \mu_{MLE}$$

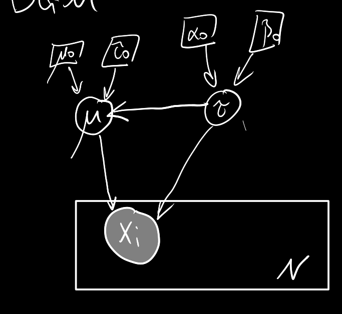
$$\sim \tau^{\frac{N-1}{2}} \exp\left(-\frac{1}{2} \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})\, \tau\right) \cdot \tau^{1/2} \exp\left(-\frac{\tau \cdot N}{2} (\bar{x} - \mu)^2\right)$$

$$p(\mu,\tau) \sim \overbrace{\underbrace{\tau^{\alpha_0 - 1} e^{-\beta_0 \tau}}_{\text{unnormalized Gamma}} \underbrace{\sqrt{\tau} \exp\left(-\frac{1}{2}(\tau_0 \tau) \cdot (\mu_0 - \mu)^2\right)}_{\text{augmented unnormalized Normal}}}^{\text{Gauss-Gamma / Normal Gamma}}$$

$$p(\mu,\tau) = \text{NormalGamma}(\mu, \tau; \alpha_0, \beta_0, \mu_0, \tau_0)$$

$$= \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \tau^{\alpha_0 - 1} e^{-\beta_0 \tau} \cdot \sqrt{\frac{\tau_0 \tau}{2\pi}} \exp\left(-\frac{1}{2}(\tau_0 \tau)(\mu_0 - \mu)^2\right)$$

DGM



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

→ not a Normal times a Gamma

$$\tau \sim \text{Gamma}(\tau; \alpha_0, \beta_0)$$

$$\mu \sim \text{Normal}(\mu | \mu_0, (\tau_0 \tau)^{-1})$$

$$X_i \sim \text{Normal}(X_i | \mu, \tau)$$

joint

$$p(\tau, \mu, \underline{X} = D) = p(\tau)\, p(\mu|\tau)\, p(\underline{X} = D | \mu, \tau)$$

$$p(\tau, \mu, D) = p(\tau)\, p(\mu|\tau) \prod_{i=0}^{N-1} p(x_i | \mu, \tau)$$

$$= \cancel{\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}} \tau^{\alpha_0 - 1} e^{-\beta_0 \tau} \cdot \sqrt{\frac{\tau_0 \tau}{\cancel{2\pi}}} \exp\left(-\frac{\tau_0 \tau}{2} \cdot (\mu - \mu_0)^2\right)$$

$$\cdot \left(\sqrt{\frac{\tau}{\cancel{2\pi}}}\right)^N \exp\left(-\frac{\tau}{2} \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)$$

posterior

Bayes' Rule

$$p(\tau, \mu | D) \stackrel{\text{Bayes' Rule}}{=} \frac{p(D|\tau,\mu)\, p(\tau,\mu)}{p(D)} \sim p(D|\tau,\mu) \cdot p(\tau,\mu)$$

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

$$p(\tau, \mu | D) \sim \tau^{\alpha_0 - 1} e^{-\beta_0 \tau} \cdot \tau^{1/2} \exp\left(-\frac{\tau_0 \tau}{2} (\mu - \mu_0)^2\right)$$

$$\cdot\, \tau^{N/2} \cdot \exp\left(-\frac{\tau}{2} \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)$$

$$= \tau^{\alpha_0 + \frac{N}{2} - 1} e^{-\beta_0 \tau}\, \tau^{1/2} \exp\left(-\frac{\tau_0 \tau}{2} (\mu - \mu_0)^2 - \frac{\tau}{2} \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)$$

$$= \tau^{\alpha_0 + \frac{N}{2} - 1} e^{-\beta_0 \tau}\, \tau^{1/2} \exp\left(-\frac{\tau}{2} \underbrace{\left(\tau_0 (\mu - \mu_0)^2 + \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)}_{=:\xi}\right)$$

$$\xi = \tau_0 (\mu - \mu_0)^2 + \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2$$

$$= \underset{a}{\tau_0} (\mu - \underset{y}{\mu_0})^2 + \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2 + \underset{b}{N} (\underset{z}{\bar{x}} - \mu)^2$$

$$\bar{x} = \frac{1}{N} \sum_{i=0}^{N-1} x^{[i]}$$

→ Sum of two quadratic forms

$a = \tau_0 \quad y = \mu_0 \quad b = N \quad z = \bar{x}$

$$c = a + b = \tau_0 + N$$

$$d = \frac{ay + bz}{a + b} = \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}$$

$$e = ab \cdot \frac{(z - y)^2}{a + b} = \tau_0 N \cdot \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N}$$

$$\xi = c \cdot (\mu - d)^2 + e + \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2$$

$$= (\tau_0 + N) \left(\mu - \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}\right)^2 + \tau_0 N \cdot \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N}$$

$$+ \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2$$

$$p(\tau, \mu | D) \sim \tau^{\alpha_0 + \frac{N}{2} - 1} e^{-\beta_0 \tau}\, \tau^{1/2} \cdot \exp\left(-\frac{\tau}{2} \xi\right)$$

$$= \tau^{\alpha_0 + \frac{N}{2} - 1} e^{-\beta_0 \tau}\, \tau^{1/2} \cdot \exp\left(-\frac{\tau}{2} \cdot (\tau_0 + N) \left(\mu - \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}\right)^2\right)$$

$$\cdot \exp\left(-\frac{\tau}{2} \cdot \tau_0 N \cdot \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N}\right) \cdot \exp\left(-\frac{\tau}{2} \cdot \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2\right)$$

$$= \tau^{\overbrace{\alpha_0 + \frac{N}{2}}^{\alpha_N} - 1} \cdot \exp\left(-\left(\overbrace{\beta_0 + \frac{1}{2} \tau_0 N \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \frac{1}{2} \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2}^{\beta_N}\right) \tau\right)$$

$$\tau^{1/2} \exp\left(-\frac{\tau}{2} \cdot \underbrace{(\tau_0 + N)}_{\tau_N} \left(\mu - \underbrace{\frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}}_{\mu_N}\right)^2\right)$$

just another Gauss-Gamma

$$\boxed{\alpha_N = \alpha_0 + \frac{N}{2}}$$

$$\boxed{\beta_N = \beta_0 + \frac{1}{2} \tau_0 N \cdot \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \frac{1}{2} \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2}$$

$$\boxed{\tau_N = \tau_0 + N}$$

$$\boxed{\mu_N = \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}}$$

$$p(\mu, \tau | D) = \text{Normal Gamma}(\mu, \tau; \alpha_N, \beta_N, \mu_N, \tau_N)$$

$$= \frac{\beta_N^{\alpha_N}}{\Gamma(\alpha_N)} \tau^{\alpha_N - 1} e^{-\beta_N \tau} \cdot \sqrt{\frac{\tau \cdot \tau_N}{2\pi}} \cdot$$

$$\cdot \exp\left(-\frac{(\tau \tau_0)}{2} \cdot (\mu - \mu_N)^2\right)$$

MAP

→ mode of the Normal-Gamma

$$\mu_{MAP}, \tau_{MAP} = \underset{\mu,\ \tau > 0}{\arg\max} \left(p(\mu, \tau | D) \right)$$

$$\log p(\mu, \tau | D) \stackrel{+}{=} (\alpha_N - 1) \log(\tau) - \beta_N \tau$$

$$+ \frac{1}{2} \log(\tau) - \frac{\tau \tau_N}{2} (\mu - \mu_N)^2$$

$$\frac{\partial \log p}{\partial \mu} = \cancel{-\frac{\tau \cdot \tau_N}{2} \cdot 2} (\mu - \mu_N) \stackrel{!}{=} 0$$

$$\mu - \mu_N = 0$$

$$\rightarrow \boxed{\mu_{MAP} = \mu_N}$$

$$\frac{\partial \log p}{\partial \tau} = \frac{\alpha_N - 1}{\tau} - \beta_N + \frac{1}{2\tau} \cancel{- \frac{1}{2} \tau_N (\mu - \mu_N)} \stackrel{!}{=} 0 \quad | \cdot \tau$$

$\cancel{\mu_N - \mu_N}$

$$\alpha_N - 1 - \beta_N \tau + \frac{1}{2} = 0$$

$$\alpha_N - \frac{1}{2} - \beta_N \tau = 0$$

$$\Leftrightarrow \boxed{\tau_{MAP} = \frac{\alpha_N - \frac{1}{2}}{\beta_N}}$$

with values plugged in

$$\boxed{\mu_{MAP} = \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}}$$

$\bar{x} \mathrel{\hat{=}} \mu_{MLE}$

→ a linear combination

$$\boxed{\tau_{MAP} = \frac{\alpha_0 + \frac{N-1}{2}}{\beta_0 + \frac{1}{2} N \cdot \tau_0 \cdot \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \frac{1}{2} \sum_{i=0}^{N-1} (x^{(i)} - \bar{x})^2}}$$

$$\tau := \frac{1}{\sigma^2}$$

$$\boxed{\sigma_{MAP} = \sqrt{\frac{\beta_0 + \frac{1}{2} N \cdot \tau_0 \cdot \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \frac{1}{2} \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2}{\alpha_0 + \frac{N-1}{2}}}}$$

$\bar{x} = \mu_{MLE}$

$\sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2 = N \cdot \sigma^2_{MLE}$