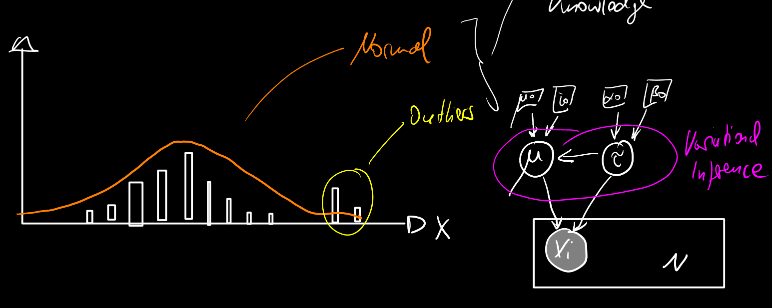
VI by Mean Field Approach for Normal/Gaussian with unknown mean & unknown precision

→ Estimate with prior knowledge



posterior $p(\mu, \tau | D)$ is tractable (here) but assume it wasn't

$$\tau \sim \text{Gamma}(\tau; \alpha_0, \beta_0)$$
$$\mu \sim \text{Normal}(\mu; \mu_0, (\tau \cdot \tau_0)^{-1})$$
$$X_i \sim \text{Normal}(X_i | \mu, \tau^{-1})$$

Variational Inference:

$$q(\mu, \tau) \approx p(\mu, \tau | D)$$

Mean Field Approach: $q(\mu, \tau) = q(\mu)\, q(\tau)$

$$\log q(\mu) = \mathbb{E}_{\tau \sim q(\tau)}[\log p(\mu, \tau, D)]$$

$\underline{X} = D$

$$\log q(\tau) = \mathbb{E}_{\mu \sim q(\mu)}[\underbrace{\log p(\mu, \tau, D)}_{\text{find log joint}}]$$

joint

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

fix $\underline{X}$ to D

$$p(\mu, \tau, \underline{X} = D) \overset{\text{i.i.d.}}{=} p(\tau)\, p(\mu|\tau) \cdot \prod_{i=0}^{N-1} p(X_i = x^{[i]} | \mu, \tau)$$

$$= \frac{\cancel{\beta_0^{\alpha_0}}}{\cancel{\Gamma(\alpha_0)}} \cdot \tau^{\alpha_0 - 1} e^{-\beta_0 \tau} \sqrt{\frac{\tau \tau_0}{\cancel{2\pi}}} \exp\left(-\frac{1}{2} \cdot (\tau \cdot \tau_0) \cdot (\mu - \mu_0)^2\right)$$
$$\cdot \left(\sqrt{\frac{\tau}{\cancel{2\pi}}}\right)^N \cdot \exp\left(-\frac{\tau}{2} \cdot \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)$$

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

$$\log p(\mu, \tau, D) \overset{+}{=} (\alpha_0 - 1) \log \tau - \beta_0 \tau + \frac{1}{2} \log \tau$$
$$- \frac{1}{2}(\tau \tau_0) \cdot (\mu - \mu_0)^2 + \frac{N}{2} \log \tau - \frac{\tau}{2} \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2$$

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

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

Sum of squared differences from the mean

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

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

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

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

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

$$\log p(\mu, \tau, D) \overset{+}{=} \left(\alpha_0 + \frac{1}{2} + \frac{N}{2} - 1\right) \log \tau - \beta_0 \tau$$
$$- \frac{\tau}{2} \cdot \left((\tau_0 + N)\left(\mu - \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}\right)^2 + \tau_0 N \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2\right)$$

Derive $q(\mu)$ & $q(\tau)$

$$\log q(\mu) = \mathbb{E}_{\tau \sim q(\tau)}[\log p(\mu, \tau, D)]$$

$$\overset{+}{=} \mathbb{E}_{\tau \sim q(\tau)}\Big[\cancel{\left(\alpha_0 + \frac{1}{2} + \frac{N}{2} - 1\right) \log \tau} \cancel{- \beta_0 \tau}$$
$$- \frac{\tau}{2} \cdot \left((\tau_0 + N)\left(\mu - \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}\right)^2 + \cancel{\tau_0 N \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N}} + \cancel{\sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2}\right)\Big]$$

$$\overset{+}{=} \mathbb{E}_{\tau}\left[-\frac{\tau}{2}(\tau_0 + N) \cdot \left(\mu - \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}\right)^2\right]$$

$$= -\frac{\tau_0 + N}{2} \cdot \mathbb{E}_{\tau}[\tau]\left(\mu - \frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}\right)^2 \quad | \exp(\ldots)$$

$$q(\mu) \propto \exp\left(-\frac{\overbrace{(\tau_0 + N) \cdot \mathbb{E}_{\tau}[\tau]}^{\tau_S}}{2}\left(\mu - \underbrace{\frac{\tau_0 \mu_0 + N\bar{x}}{\tau_0 + N}}_{\mu_S}\right)^2\right)$$

↳ a Normal distribution

$$q(\mu) = \mathcal{N}(\mu; \mu_S, \tau_S^{-1})$$

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

$$\boxed{\tau_S = (\tau_0 + N)\, \mathbb{E}_{\tau}[\tau]}$$

figure it out shortly

---

now $q(\tau)$

$$\log q(\tau) = \mathbb{E}_{\mu \sim q(\mu)}[\log p(\mu, \tau, D)]$$

$$= \mathbb{E}_{\mu}\Big[\left(\alpha_0 + \frac{N+1}{2} - 1\right) \log \tau$$
$$- \left(\beta_0 + \frac{1}{2}(\tau_0 + N) \cdot (\mu - \mu_S)^2 + \frac{\tau_0 N}{2} \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N}\right.$$
$$\left. + \frac{1}{2} \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2\right)\tau\Big]$$

$$= \left(\overbrace{\alpha_0 + \frac{N+1}{2}}^{\alpha_S} - 1\right) \log \tau$$
$$- \overbrace{\left(\beta_0 + \frac{1}{2}(\tau_0 + N)\, \mathbb{E}_{\mu}[(\mu - \mu_S)^2] + \frac{\tau_0 N}{2} \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \frac{1}{2} \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2\right)}^{\beta_S} \tau$$

→ Gamma distribution

$$q(\tau) = \text{Gamma}(\tau; \alpha_S, \beta_S)$$

$$\boxed{\alpha_S = \alpha_0 + \frac{N+1}{2}}$$

$$\boxed{\beta_S = \beta_0 + \frac{1}{2}(\tau_0 + N) \cdot \mathbb{E}_{\mu}[(\mu - \mu_S)^2] + \frac{\tau_0 N}{2} \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \frac{1}{2} \sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2}$$

Recap

$$q(\mu) = \mathcal{N}(\mu; \mu_S, \tau_S^{-1})$$
$$q(\tau) = \text{Gamma}(\tau; \alpha_S, \beta_S)$$

still contains Expectation wrt. two other distributions

→ solve by analytical forms of the moments

$$\mathbb{E}_{\tau \sim \text{Gamma}(\tau; \alpha_S, \beta_S)}[\tau] = \frac{\alpha_S}{\beta_S}$$

$$\mathbb{E}_{\mu \sim \mathcal{N}(\mu; \mu_S, \tau_S^{-1})}[(\mu - \mu_S)^2] = \tau_S^{-1}$$

in total

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

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

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

$$\boxed{\alpha_S = \alpha_0 + \frac{N+1}{2}}$$

$$\boxed{\tau_S = (\tau_0 + N) \frac{\alpha_S}{\beta_S}}$$

$$\boxed{\beta_S = \beta_0 + \frac{1}{2}(\tau_0 + N) \cdot \tau_S^{-1} + \frac{\tau_0 N}{2} \frac{(\mu_0 - \bar{x})^2}{\tau_0 + N} + \frac{1}{2} \underbrace{\cancel{\sum_{i=0}^{N-1} (x^{[i]} - \bar{x})^2}}_{N \sigma^2_{MLE}}}$$

Circular dependency between τ_S & β_S

⇒ iterative algorithm

→ Surrogate MAP estimate by Maximizing the Surrogate

$$\boxed{\mu_{MAP,S} = \mu_S}$$

$$\boxed{\tau_{MAP,S} = \frac{\alpha_S - 1}{\beta_S}}$$