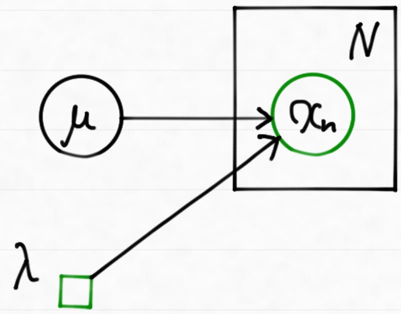
§3.3 1次元Gauss分布の学習と予測.

精度パラメータ : $\lambda := \frac{1}{\sigma^2}$ とおく.

3.3.1. 平均が未知の場合



・ $\lambda \in \mathbb{R}_{>0}$: fixed. μ を推定したい.

$$x \sim p(x|\mu) = \mathcal{N}(x|\mu, \lambda^{-1}).$$

μ に対しては、共役事前分布 は Gauss分布.

$$p(\mu) = \mathcal{N}(\mu|m, \lambda_\mu^{-1})$$ m, λ_μ : ハイパーパラメータ.

・ データ $\mathcal{X} = \{x_1, \cdots, x_N\}$: given.

$$p(\mu|\mathcal{X}) \propto p(\mathcal{X}|\mu)p(\mu) \quad (\because \text{Bayes' Th.})$$
$$= \left(\prod_{n=1}^{N} p(x_n|\mu)\right) p(\mu)$$
$$= \left(\prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \lambda^{-1})\right) \mathcal{N}(\mu|m, \lambda_\mu^{-1}).$$

$$\log p(\mu|\mathcal{X}) = \sum_{n=1}^{N} \log \mathcal{N}(x_n|\mu, \lambda^{-1}) + \log \mathcal{N}(\mu|m, \lambda_\mu^{-1}) + \text{const}.$$

$$= \sum_{n=1}^{N} \left(\underbrace{\frac{1}{2}\log\lambda - \frac{1}{2}\log 2\pi}_{\text{const.}} - \frac{1}{2}\lambda(x_n - \mu)^2 \right)$$

今興味あるのは μ の項のみ.

$$+ \underbrace{\frac{1}{2}\log\lambda_\mu - \frac{1}{2}\log 2\pi}_{\text{const.}} - \frac{1}{2}\lambda_\mu(\mu - m)^2 + \text{const}.$$

$$= -\frac{1}{2}\sum_{n=1}^{N}(\underbrace{\lambda x_n^2}_{\text{const.}} - 2\lambda x_n\mu + \lambda\mu^2) - \frac{1}{2}\lambda_\mu\mu^2 + \lambda_\mu m\mu \underbrace{- \frac{1}{2}\lambda_\mu m^2}_{\text{const.}} + \text{const}.$$

$$= -\frac{1}{2}\left((N\lambda + \lambda_\mu)\mu^2 - 2\left(\lambda\sum_{n=1}^{N} x_n + m\lambda_\mu\right)\mu \right) + \text{const}.$$

→ μ に関する上に凸の2次関数なので μ の事後分布は Gauss分布.

$\rightarrow p(\mu|\mathcal{X}) = \mathcal{N}(\mu|\hat{m}, \hat{\lambda}_\mu^{-1})$ と書けるとする.

$$\log p(\mu|\mathcal{X}) = \frac{1}{2}\log\hat{\lambda}_\mu - \frac{1}{2}\log 2\pi - \frac{\hat{\lambda}_\mu}{2}(\mu-\hat{m})^2$$

$$= -\frac{1}{2}(\hat{\lambda}_\mu\mu^2 - 2\hat{m}\hat{\lambda}_\mu\mu) + \mathrm{const.}$$

先の式

$$\log p(\mu|\mathcal{X}) = -\frac{1}{2}\left(\underbrace{(N\lambda+\lambda_\mu)}_{=:\hat{\lambda}_\mu}\mu^2 - 2\underbrace{\left(\lambda\sum_{n=1}^{N}x_n + m\lambda_\mu\right)}_{=:\hat{m}\hat{\lambda}_\mu}\mu\right) + \mathrm{const.}$$

と比較して,

$$\hat{\lambda}_\mu = N\lambda+\lambda_\mu, \quad \hat{m} = \frac{1}{\hat{\lambda}_\mu}\left(\lambda\sum_{n=1}^{N}x_n + m\lambda_\mu\right)$$

と求まる.

○ 解釈

・事後分布の精度: $\hat{\lambda}_\mu = N\lambda + \lambda_\mu \xrightarrow{N\to\infty} \infty$.

$\rightarrow$ μに対する精度がデータの観測により上昇していく.

・事後分布の平均: $\hat{m} = \frac{1}{\hat{\lambda}_\mu}\left(\lambda\sum_{n=1}^{N}x_n + m\lambda_\mu\right)$

$$= \frac{1}{N\lambda+\lambda_\mu}\left(N\lambda\left(\frac{1}{N}\sum_{n=1}^{N}x_n\right) + \lambda_\mu m\right)$$

データ平均 $\frac{1}{N}\sum x_n$ と m の加重平均

$\rightarrow$ $N\to\infty$ でデータ平均の影響が大きくなっていく.

- 未観測データに対する予測分布.

- $$p(x_*) = \int p(x_* | \mu) p(\mu) d\mu$$
$$= \int \mathcal{N}(x_* | \mu, \lambda^{-1}) \mathcal{N}(\mu | m, \lambda_\mu^{-1}) d\mu.$$

これを直接計算するのは面倒. 対数を使う.

- Bayes' Thm. より.

$$p(\mu | x_*) = \frac{p(x_* | \mu) p(\mu)}{p(x_*)}$$

この対数をとると

$$\log p(\mu | x_*) = \log p(x_* | \mu) - \log p(x_*) + \underline{\log p(\mu)}.$$

← x_* に依らないので const. として扱えばよい.

$$\therefore \log p(x_*) = \log p(x_* | \mu) - \log p(\mu | x_*) + \text{const}.$$

- $p(\mu | x_*)$ について.

x_* : given のときの μ の条件付き分布とみると，事後分布の計算と同様.

→ N コあるデータ X を，1コのデータ x_* におきかえて考えると，

$$p(\mu | x_*) = \mathcal{N}\left(\mu \,\middle|\, \underbrace{\frac{\lambda x_* + \lambda_\mu m}{\lambda + \lambda_\mu}}_{=: m(x_*)}, (\lambda + \lambda_\mu)^{-1}\right)$$

$$\log p(\mu | x_*) = -\frac{1}{2}(\lambda + \lambda_\mu)(\mu - m(x_*))^2 + \text{const}.$$
$$= -\frac{1}{2}(\lambda + \lambda_\mu)(m(x_*)^2 - 2\mu m(x_*)) + \text{const}.$$

$\therefore \log p(x_*)$

$$= -\frac{1}{2}\left(\lambda(x_*^2 - 2\mu x_*) - (\lambda+\lambda_\mu)\left(m(x_*)^2 - 2\mu m(x_*)\right)\right) + \text{const.}$$

$$= -\frac{1}{2}\left(\lambda x_*^2 - \cancel{2\mu\lambda x_*} - \frac{(\lambda x_* + \lambda_\mu m)^2}{\lambda+\lambda_\mu} + 2\mu\left(\cancel{\lambda x_*} + \underbrace{\lambda_\mu m}_{\text{const.}}\right)\right) + \text{const.}$$

$$= -\frac{1}{2}\left(\lambda x_*^2 - \frac{1}{\lambda+\lambda_\mu}\left(\lambda^2 x_*^2 + 2\lambda\lambda_\mu m x_* + \underbrace{\lambda_\mu^2 m^2}_{\text{const.}}\right)\right) + \text{const.}$$

$$= -\frac{1}{2}\left(\underbrace{\frac{\lambda\lambda_\mu}{\lambda+\lambda_\mu}}_{=:\lambda_*} x_*^2 - 2m\underbrace{\frac{\lambda\lambda_\mu}{\lambda+\lambda_\mu}}_{=:\mu_*\lambda_*} x_*\right) + \text{const.}$$

← Gauss分布の対数表示.

$\therefore p(x_*) = \mathcal{N}(x_* \mid \mu_*, \lambda_*^{-1}),$

$$\lambda_* = \frac{\lambda\lambda_\mu}{\lambda+\lambda_\mu}, \quad \mu_* = m.$$

↓

分散で考えると、$\lambda_*^{-1} = \lambda^{-1} + \lambda_\mu^{-1}$. つまり、

(予測分布の不確かさ) = (観測分布の不確かさ)

+ (事前分布の不確かさ).

- $p(x_* \mid \mathcal{X})$ を求めるときは、上の式の m, λ_μ を $\hat{m}, \hat{\lambda}_\mu$ にする.
- λ の分布も学習したいときは、ガンマ事前分布をモデルに追加して推論する.

3.3.2 精度が未知の場合.

μ: given, λ: unknown. → λを推定.

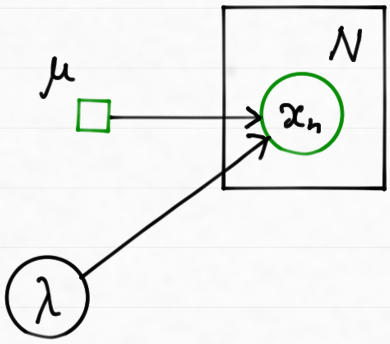
◦観測モデル: $p(x|\lambda) = \mathcal{N}(x|\mu, \lambda^{-1})$.

λの事前分布として, $\lambda \in \mathbb{R}_{>0}$ なのでガンマ事前分布を使う:

$$p(\lambda) = \mathrm{Gam}(\lambda|a, b).$$

ガンマ分布は, Gauss分布の精度パラメータに対する共役事前分布.

* パラメータの更新:



Bayes' Thm.

$$p(\lambda|\mathcal{X}) \propto p(\mathcal{X}|\lambda)p(\lambda)$$
$$= \left(\prod_{n=1}^{N} p(x_n|\lambda)\right)p(\lambda)$$
$$= \left(\prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \lambda^{-1})\right)\mathrm{Gam}(\lambda|a, b).$$

この対数をとると,

$$\log p(\lambda|\mathcal{X}) = \sum_{n=1}^{N} \log \mathcal{N}(x_n|\mu, \lambda^{-1}) + \log \mathrm{Gam}(\lambda|a, b) + \mathrm{const.}$$

$$= \sum_{n=1}^{N}\left(-\frac{1}{2}\left(\lambda(x_n-\mu)^2 - \log\lambda\right)\right) + (a-1)\log\lambda - b\lambda + \mathrm{const.}$$

$$= \left(\frac{N}{2} + a - 1\right)\log\lambda - \left(\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2 + b\right)\lambda + \mathrm{const.}$$

→ ガンマ分布のpdfに対数をとったものと形が同じ.

$\therefore\ p(\lambda|\mathbf{x}) = \mathrm{Gam}(\lambda|\hat{a},\hat{b}),$

$$\hat{a} := \frac{N}{2}+a,\quad \hat{b} := \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2+b$$

以上より共役事前分布がガンマ分布であることが確認できた．

* 予測分布の計算．

$p(x_*) = \int p(x_*|\lambda)p(\lambda)d\lambda$ を計算してもよいが面倒．対数を活用する

$$p(\lambda|x_*) = \frac{p(x_*|\lambda)p(\lambda)}{p(x_*)} \quad (\because \text{ Bayes' Thm.})$$

対数をとり整理

$$\therefore\ \log p(x_*) = \log p(x_*|\lambda) - \log p(\lambda|x_*) + \log p(\lambda).$$

・第1項はモデルの式より，

$$\log p(x_*|\lambda) = \log \mathcal{N}(x_*|\mu,\lambda^{-1}) = -\frac{1}{2}\left(\lambda(x_*-\mu)^2 - \log\lambda\right) + \mathrm{const}$$

・第2項．

$p(\lambda|x_*)$を「x_*：givenのときの事後分布」とみれば，先の計算より

$$p(\lambda|x_*) = \mathrm{Gam}\left(\lambda\middle|\frac{1}{2}+a,\, b(x_*)\right),\quad b(x_*) := \frac{1}{2}(x_*-\mu)^2+b.$$

$\therefore\ \log p(\lambda|x_*)$

ほしいのは x_* についての項．
λについては const. であっても捨ててはいけない．

$$= \left(\frac{1}{2}+a-1\right)\log\lambda - \left(\frac{1}{2}(x_*-\mu)^2+b\right)\lambda + \underline{\log C_G(a, b(x_*))}$$

$$= \left(\frac{1}{2}+a-1\right)\log\lambda - \left(\frac{1}{2}(x_*-\mu)^2+b\right)\lambda + \left(\frac{1}{2}+a\right)\log\left(\frac{1}{2}(x_*-\mu)^2+b\right) + \mathrm{const}$$

・第3項は事前分布の式より $\log p(\lambda) = (a-1)\log\lambda - b\lambda + \mathrm{const}.$

$$\therefore \log p(x_*) = \log p(x_*|\lambda) - \log p(\lambda|x_*) + \log p(\lambda)$$

$$= -\frac{1}{2}\left(\cancel{\lambda(x_*-\mu)^2} - \cancel{\log\lambda}\right) - \cancel{\left(\frac{1}{2}+a-1\right)\log\lambda} + \left(\cancel{\frac{1}{2}(x_*-\mu)^2} + \cancel{b}\right)\lambda$$

$$- \left(\frac{1}{2}+a\right)\log\left(\frac{1}{2}(x_*-\mu)^2 + b\right) + \cancel{(a-1)\log\lambda} - \cancel{b\lambda} + \text{const.}$$

$$= -\frac{2a+1}{2}\log\left(b\left(1+\frac{1}{2b}(x_*-\mu)^2\right)\right) + \text{const.}$$

$$= -\frac{2a+1}{2}\log\left(1+\frac{1}{2b}(x_*-\mu)^2\right) + \text{const.}$$

→実は、これはStudentのt分布のpdfの対数をとったものと同じ形.

Def. (Studentのt分布) $\mu_s \in \mathbb{R}, \lambda_s, \nu_s \in \mathbb{R}_{>0}$.

$$\text{pdf.}: \mathrm{St}(x|\mu_s, \lambda_s, \nu_s) = \frac{\Gamma\left(\frac{\nu_s+1}{2}\right)}{\Gamma\left(\frac{\nu_s}{2}\right)}\left(\frac{\lambda_s}{\pi\nu_s}\right)^{\frac{1}{2}}\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}}.$$

$(x \in \mathbb{R})$ □

- $\log \mathrm{St}(x|\mu_s, \lambda_s, \nu_s) = -\frac{\nu_s+1}{2}\log\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right) + \text{const.}$

- 以上より, $p(x_*) = \mathrm{St}(x_*|\mu_s, \lambda_s, \nu_s)$,

$$\mu_s = \mu, \quad \lambda_s = \frac{a}{b}, \quad \nu_s = 2a.$$

cf. Studentのt分布の正規化項, 平均, 分散の計算.

◦ 正規化項が $\frac{\Gamma\left(\frac{\nu_s+1}{2}\right)}{\Gamma\left(\frac{\nu_s}{2}\right)}\left(\frac{\lambda_s}{\pi\nu_s}\right)^{\frac{1}{2}}$ になるのを確認する.

$$I = \int_{-\infty}^{\infty}\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}}dx = \frac{\Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s+1}{2}\right)}\left(\frac{\pi\nu_s}{\lambda_s}\right)^{\frac{1}{2}}$$ を示せばよい.

まず, $\sqrt{\frac{\lambda_s}{\nu_s}}(x-\mu_s) = t$ とおくと, $I = \sqrt{\frac{\nu_s}{\lambda_s}}\int_{-\infty}^{\infty}(1+t^2)^{-\frac{\nu_s+1}{2}}dt.$

被積分関数はtの偶関数なので、

$$I = 2\sqrt{\frac{\nu_s}{\lambda_s}}\int_0^{\infty}\left(1+t^2\right)^{-\frac{\nu_s+1}{2}}dt.$$

$u = (1+t^2)^{-1}$ とおくと、 $t = \sqrt{\frac{1-u}{u}}$,

$$du = -(1+t^2)^{-2}\cdot 2t\,dt = -2u^2\sqrt{\frac{1-u}{u}}\,dt.$$

$$\therefore dt = -\frac{1}{2}u^{-\frac{3}{2}}(1-u)^{-\frac{1}{2}}du.$$

積分範囲は $0\to\infty$ が $1\to 0$ に変わる。

$$\therefore I = \sqrt{\frac{\nu_s}{\lambda_s}}\int_0^1 u^{\frac{\nu_s+1}{2}}\cdot u^{-\frac{3}{2}}(1-u)^{-\frac{1}{2}}du$$

$$= \sqrt{\frac{\nu_s}{\lambda_s}}\int_0^1 u^{\frac{\nu_s}{2}-1}(1-u)^{\frac{1}{2}-1}du$$

ベータ関数の定義より

$$= \sqrt{\frac{\nu_s}{\lambda_s}}\,B\left(\frac{\nu_s}{2},\frac{1}{2}\right)$$

ベータ関数の性質 $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

$$= \sqrt{\frac{\nu_s}{\lambda_s}}\frac{\Gamma\left(\frac{\nu_s}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{\nu_s+1}{2}\right)}$$

$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

$$= \frac{\Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s+1}{2}\right)}\left(\frac{\pi\nu_s}{\lambda_s}\right)^{\frac{1}{2}}.$$

□

- 平均の計算。

$C_t(\mu_s,\lambda_s,\nu_s) := \frac{\Gamma\left(\frac{\nu_s+1}{2}\right)}{\Gamma\left(\frac{\nu_s}{2}\right)}\left(\frac{\lambda_s}{\pi\nu_s}\right)^{\frac{1}{2}}$ とおく。

$$\mathbb{E}[X] = C_t(\mu_s,\lambda_s,\nu_s)\int_{-\infty}^{\infty}x\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}}dx$$

$$= C_t(\mu_s, \lambda_s, \nu_s) \int_{-\infty}^{\infty} (x-\mu_s)\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}} dx$$

$$+ \mu_s \underbrace{\int_{-\infty}^{\infty} C_t(\mu_s, \lambda_s, \nu_s)\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}} dx}_{= \int \mathrm{St}(x|\mu_s, \lambda_s, \nu_s)}.$$

$$= \mu_s + C_t(\mu_s, \lambda_s, \nu_s) \int_{-\infty}^{\infty} (x-\mu_s)\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}} dx.$$

$$\int_{-\infty}^{\infty} (x-\mu_s)\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}} dx \quad \Big\downarrow \ \sqrt{\frac{\lambda_s}{\nu_s}}(x-\mu_s) = t \text{ とおく}$$

$$= \frac{\nu_s}{\lambda_s} \int_{-\infty}^{\infty} t(1+t^2)^{-\frac{\nu_s+1}{2}} dt$$

$$= \frac{\nu_s}{\lambda_s}\left(\int_0^{\infty} t(1+t^2)^{-\frac{\nu_s+1}{2}} dt + \int_{-\infty}^{0} t(1+t^2)^{-\frac{\nu_s+1}{2}} dt \right) \quad \Big\downarrow \ 1+t^2 = u \text{ とおく}$$

$$= \frac{\nu_s}{2\lambda_s}\left(\int_1^{\infty} u^{-\frac{\nu_s+1}{2}} du + \int_{\infty}^{1} u^{-\frac{\nu_s+1}{2}} du \right).$$

ここで，$\nu_s \neq 1$ のとき，

$$\int_1^{\infty} u^{-\frac{\nu_s+1}{2}} du = \lim_{a\to\infty} \int_1^{a} u^{-\frac{\nu_s+1}{2}} du$$

$$= \lim_{a\to\infty} \left[-\frac{2}{\nu_s-1} u^{-\frac{\nu_s-1}{2}} \right]_1^a$$

$$= \lim_{a\to\infty} \frac{2}{\nu_s-1}\left(1-a^{-\frac{\nu_s-1}{2}}\right)$$

$$= \begin{cases} \dfrac{2}{\nu_s-1} & (\nu_s > 1) \\ -\infty & (0 < \nu_s < 1) \end{cases}.$$

$\nu_s = 1$ では,

$$\int_1^\infty u^{-\frac{\nu_s+1}{2}} du = \int_1^\infty u^{-1} du = \lim_{a\to\infty} \left[\log u\right]_1^a = \infty.$$

同様に,

$$\int_\infty^1 u^{-\frac{\nu_s+1}{2}} du = \begin{cases} -\dfrac{2}{\nu_s-1} & (\nu_s > 1) \\ -\infty & (\nu_s = 1) \\ \infty & (0 < \nu_s < 1) \end{cases}.$$

以上より,

$$\int_1^\infty u^{-\frac{\nu_s+1}{2}} du + \int_\infty^1 u^{-\frac{\nu_s+1}{2}} du = \begin{cases} 0 & (\nu_s > 1) \\ \text{indeterminate} & (0 < \nu_s \leq 1) \end{cases}$$

$$\therefore \mathbb{E}[X] = \begin{cases} \mu_s & (\nu_s > 1) \\ \text{undefined} & (0 < \nu_s \leq 1) \end{cases}.$$

- 分散の計算.

$\mathbb{E}[X]$ は $\nu_s > 1$ のとき存在するので, $\nu_s > 1$ の範囲で考える.

$$\mathbb{V}[X]$$

$$= \mathbb{E}[(X-\mu_s)^2]$$

$$= C_t(\mu_s, \lambda_s, \nu_s) \int_{-\infty}^{\infty} (x-\mu_s)^2 \left(1 + \frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}} dx.$$

$$\int_{-\infty}^{\infty}(x-\mu_s)^2\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}}dx$$

$\downarrow \sqrt{\frac{\lambda_s}{\nu_s}}(x-\mu_s)=t$ とおく

$$=\left(\frac{\nu_s}{\lambda_s}\right)^{\frac{3}{2}}\int_{-\infty}^{\infty}t^2(1+t^2)^{-\frac{\nu_s+1}{2}}dt$$

$\downarrow$ tについて偶関数

$$=\left(\frac{\nu_s}{\lambda_s}\right)^{\frac{3}{2}}\cdot 2\int_{0}^{\infty}t^2(1+t^2)^{-\frac{\nu_s+1}{2}}dt$$

ここで，$(1+t^2)^{-1}=u$ とおくと，$t\to\infty$ で $u\to 0$.

$dt=-\frac{1}{2tu^2}du$ で，$t=\sqrt{\frac{1-u}{u}}$.

$$\therefore \int_{0}^{\infty}t^2(1+t^2)^{-\frac{\nu_s+1}{2}}dt$$

$$=\int_{1}^{0}\sqrt{\frac{1-u}{u}}\cdot u^{\frac{\nu_s+1}{2}}\left(-\frac{1}{2u^2}\right)du$$

$$=\frac{1}{2}\int_{0}^{1}u^{\frac{\nu_s-2}{2}-1}(1-u)^{\frac{3}{2}-1}du$$

$$=\begin{cases}\frac{1}{2}B\left(\frac{\nu_s-2}{2},\frac{3}{2}\right) & (\nu_s>2)\\ \text{undefined} & (1<\nu_s\leq 2)\end{cases}$$

ベータ関数$B(x,y)$の定義域は $x>0, y>0$.

この範囲では積分は収束しない.

よって，$\nu_s>2$ の範囲で，

$$\int_{-\infty}^{\infty}(x-\mu_s)^2\left(1+\frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right)^{-\frac{\nu_s+1}{2}}dx$$

$$=\left(\frac{\nu_s}{\lambda_s}\right)^{\frac{3}{2}}B\left(\frac{\nu_s-2}{2},\frac{3}{2}\right)$$

$$=\left(\frac{\nu_s}{\lambda_s}\right)^{\frac{3}{2}}\frac{\Gamma\left(\frac{\nu_s-2}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{\nu_s+1}{2}\right)}.$$

このとき，

$$\mathbb{V}[X] = C_t(\mu_s, \lambda_s, \nu_s)\left(\frac{\nu_s}{\lambda_s}\right)^{\frac{3}{2}} \frac{\Gamma\left(\frac{\nu_s-2}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{\nu_s+1}{2}\right)}$$

$$= \frac{\cancel{\Gamma\left(\frac{\nu_s+1}{2}\right)}}{\Gamma\left(\frac{\nu_s}{2}\right)}\left(\frac{\lambda_s}{\pi\nu_s}\right)^{\frac{1}{2}}\left(\frac{\nu_s}{\lambda_s}\right)^{\frac{3}{2}} \frac{\Gamma\left(\frac{\nu_s-2}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\cancel{\Gamma\left(\frac{\nu_s+1}{2}\right)}}$$

$$= \frac{\nu_s\,\Gamma\left(\frac{\nu_s-2}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{\nu_s}{2}\right)\sqrt{\pi}}\lambda_s^{-1}$$

$\Gamma(x+1) = x\Gamma(x)$

$$= \frac{\nu_s\,\cancel{\Gamma\left(\frac{\nu_s-2}{2}\right)}\,\cancel{\tfrac{1}{2}}\,\cancel{\Gamma\left(\frac{1}{2}\right)}}{\frac{\nu_s-2}{\cancel{2}}\,\cancel{\Gamma\left(\frac{\nu_s-2}{2}\right)}\,\cancel{\sqrt{\pi}}}\lambda_s^{-1}$$

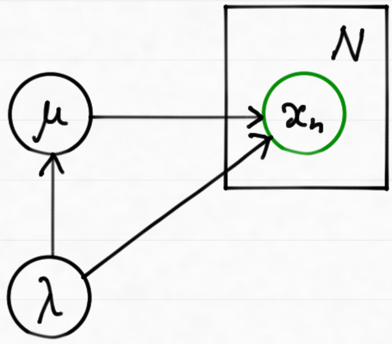
$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$$= \frac{\nu_s}{\nu_s-2}\lambda_s^{-1}.$$

$$\therefore\quad \mathbb{V}[X] = \begin{cases} \dfrac{\nu_s}{\nu_s-2}\lambda_s^{-1} & (\nu_s > 2) \\ \text{undefined} & (0 < \nu_s \leq 2). \end{cases}$$

3.3.3 平均・精度が未知の場合.

観測モデル $p(x|\mu,\lambda)=\mathcal{N}(x|\mu,\lambda^{-1})$.



共役事前分布は Gauss-ガンマ分布：

$$p(\mu,\lambda)=NG(\mu,\lambda|m,\beta,a,b)$$

$$:=\mathcal{N}(\mu|m,(\beta\lambda)^{-1})\,\mathrm{Gam}(\lambda|a,b).$$

→ データ $\mathcal{X}=\{x_1,\cdots,x_N\}$ 観測後の $\hat{m},\hat{\beta},\hat{a},\hat{b}$ を求める.

・平均 μ について.

$$\left(\begin{array}{l} p(\mu)=\mathcal{N}(\mu|m,\lambda_\mu^{-1}) \text{ とすると} \\ p(\mu|\mathcal{X})=\mathcal{N}(\mu|\hat{m},\hat{\lambda}_\mu^{-1}) \\ \quad \hat{\lambda}_\mu = N\lambda+\lambda_\mu, \quad \hat{m}=\dfrac{\lambda\sum_{n=1}^{N}x_n+\lambda_\mu m}{\hat{\lambda}_\mu} \\ \text{であった.} \end{array}\right)$$

$p(\mu|\lambda)=\mathcal{N}(\mu|m,(\beta\lambda)^{-1})$ なので,

$$p(\mu|\lambda,\mathcal{X})=\mathcal{N}(\mu|\hat{m},(\hat{\beta}\lambda)^{-1}),$$

$$\hat{\beta}=N+\beta,\quad \hat{m}=\frac{1}{\hat{\beta}}\left(\sum_{n=1}^{N}x_n+\beta m\right)$$

となる.

・精度λについて．

$p(\lambda) = \mathrm{Gam}(\lambda|a,b)$．$p(\lambda|\mathcal{X})$を求める．

$\mathcal{X}$とμとλの同時分布は，

$$p(\mathcal{X},\mu,\lambda) = p(\mu|\lambda,\mathcal{X})p(\lambda,\mathcal{X})$$
$$= p(\mu|\lambda,\mathcal{X})p(\lambda|\mathcal{X})p(\mathcal{X}).$$

よって．

$$p(\lambda|\mathcal{X}) = \frac{p(\mathcal{X},\mu,\lambda)}{p(\mu|\lambda,\mathcal{X})p(\mathcal{X})} \propto \frac{p(\mathcal{X},\mu,\lambda)}{p(\mu|\lambda,\mathcal{X})}.$$

→ $p(\mu|\lambda,\mathcal{X})$は既に求めている！$p(\mathcal{X},\mu,\lambda)$も

$$p(\mathcal{X},\mu,\lambda) = p(\mathcal{X}|\mu,\lambda)p(\mu,\lambda)$$
$$= \left(\prod_{n=1}^{N}\mathcal{N}(x_n|\mu,\lambda^{-1})\right)\mathcal{N}(\mu|m,(\beta\lambda)^{-1})\,\mathrm{Gam}(\lambda|a,b)$$

から求まる．

$\therefore$ $\log p(\lambda|\mathcal{X})$

$$= \sum_{n=1}^{N}\log\mathcal{N}(x_n|\mu,\lambda^{-1}) + \log\mathcal{N}(\mu|m,(\beta\lambda)^{-1}) + \log\mathrm{Gam}(\lambda|a,b)$$
$$- \log\mathcal{N}(\mu|\hat{m},(\hat{\beta}\lambda)^{-1}) + \mathrm{const}.$$
$$= \sum_{n=1}^{N}\left(-\frac{1}{2}\left(\lambda(x_n-\mu)^2 - \log\lambda\right)\right) - \frac{1}{2}\left(\beta\lambda(\mu-m)^2 - \log\beta\lambda\right)$$
$$+ (a-1)\log\lambda - b\lambda + \frac{1}{2}\left(\hat{\beta}\lambda(\mu-\hat{m})^2 - \log\hat{\beta}\lambda\right) + \mathrm{const}.$$

$$= \sum_{n=1}^{N} \left(-\frac{1}{2} \left(\lambda (x_n - \mu)^2 - \log \lambda \right) \right) - \frac{1}{2} \left(\beta\lambda (\mu - m)^2 - \log \beta\lambda \right)$$

$$+ (a-1)\log\lambda - b\lambda + \frac{1}{2}\left(\hat{\beta}\lambda(\mu - \hat{m})^2 - \log \hat{\beta}\lambda \right) + \text{const.}$$

$$= \left(\frac{N}{2} + a - 1 \right) \log\lambda - \frac{1}{2}\lambda \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{1}{2}\beta\lambda(\mu - m)^2$$

$$+ \frac{1}{2}\left(\underbrace{\log\beta}_{\text{const.}} + \cancel{\log\lambda} \right) - b\lambda + \frac{1}{2}\hat{\beta}\lambda(\mu - \hat{m})^2 - \frac{1}{2}\left(\underbrace{\log\hat{\beta}}_{\text{const.}} + \cancel{\log\lambda} \right) + \text{const}$$

$$= \left(\frac{N}{2} + a - 1 \right) \log\lambda - \frac{\lambda}{2}\left(\sum_{n=1}^{N} x_n^2 - \cancel{2\mu \sum_{n=1}^{N} x_n} + \cancel{N\mu^2} \right) \quad \leftarrow \hat{m}\text{と}\hat{\beta}\text{を代入する.}$$

$$- \frac{\lambda}{2}\left(\cancel{\beta\mu^2} - \cancel{2\beta\mu m} + \beta m^2 \right) - b\lambda + \frac{\lambda}{2}\left(\cancel{\hat{\beta}\mu^2} - \cancel{2\hat{\beta}\mu\hat{m}} + \hat{\beta}\hat{m}^2 \right) + \text{const}$$

$$= \left(\frac{N}{2} + a - 1 \right) \log\lambda - \left(\frac{1}{2}\left(\sum_{n=1}^{N} x_n^2 + \beta m^2 - \hat{\beta}\hat{m}^2 \right) + b \right) \lambda + \text{const.}$$

$$\therefore p(\lambda | X) = \mathrm{Gam}(\lambda | \hat{a}, \hat{b}),$$

$$\hat{a} = \frac{N}{2} + a, \quad \hat{b} = \frac{1}{2}\left(\sum_{n=1}^{N} x_n^2 + \beta m^2 - \hat{\beta}\hat{m}^2 \right) + b.$$

以上より,

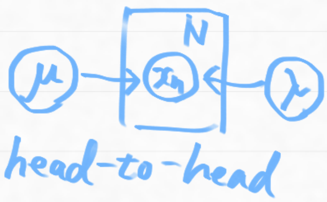
$$p(\mu, \lambda | X) = p(\mu | \lambda, X) p(\lambda | X)$$

$$= \mathcal{N}(\mu | \hat{m}, (\hat{\beta}\lambda)^{-1}) \mathrm{Gam}(\lambda | \hat{a}, \hat{b})$$

$$= \mathrm{NG}(\mu, \lambda | \hat{m}, \hat{\beta}, \hat{a}, \hat{b}).$$

- $p(\mu, \lambda) = p(\mu)p(\lambda)$ (独立なprior)を使うこともできるが、データ観測後はμ, λが独立にならないため、$p(\mu, \lambda | X)$が二つのパラメータの同時分布になる。

→ Gauss-ガンマ分布でやる方が単純.



- 予測分布の計算.

$$p(x_*) = \iint p(x_* | \mu, \lambda) p(\mu, \lambda) d\mu d\lambda.$$

→直接計算をしないで求める.

Bayes' Thm. より

$$\log p(x_*) = \log p(x_* | \mu, \lambda) - \log p(\mu, \lambda | x_*) + \log p(\mu, \lambda).$$

$$\log p(x_* | \mu, \lambda) = \log \mathcal{N}(x_* | \mu, \lambda^{-1})$$

$$= -\frac{1}{2}\left(\lambda (x_* - \mu)^2 - \log \lambda\right) + \text{const.}$$

$$\log p(\mu, \lambda | x_*) = \log p(\mu | \lambda, \{x_*\}) + \log p(\lambda | \{x_*\})$$

$$= \log \mathcal{N}\Big(\mu \Big| \underbrace{\frac{1}{1+\beta}(x_* + \beta m)}_{=: m(x_*)}, ((1+\beta)\lambda)^{-1}\Big)$$

$$+ \log \mathrm{Gam}\left(\lambda \Big| \frac{1}{2} + a, \frac{1}{2}\left(x_*^2 + \beta m^2 - (1+\beta)\left(\frac{1}{1+\beta}(x_* + \beta m)\right)^2\right) + b\right)$$

$$= \log \mathcal{N}(\mu | m(x_*), ((1+\beta)\lambda)^{-1})$$

$$+ \log \mathrm{Gam}\Big(\lambda \Big| \frac{1}{2} + a, \underbrace{\frac{\beta}{2(1+\beta)}(x_* - m)^2 + b}_{=: b(x_*)}\Big)$$

$$= \log \mathcal{N}(\mu | m(x_*), ((1+\beta)\lambda)^{-1}) + \log \mathrm{Gam}\left(\lambda \Big| \frac{1}{2} + a, b(x_*)\right)$$

$$= -\frac{1}{2}\left((1+\beta)\lambda(\mu - m(x_*))^2 - \log(1+\beta)\lambda\right)$$

$$+ \left(\frac{1}{2} + a - 1\right)\log \lambda - b(x_*)\lambda + \left(\frac{1}{2} + a\right)\log b(x_*) + \text{const.}$$

$$\log p(\mu,\lambda) = \log \mathcal{N}(\mu|m,(\beta\lambda)^{-1}) + \log \mathrm{Gam}(\lambda|a,b)$$

$$= -\frac{1}{2}\left(\beta\lambda(\mu-m)^2 - \log\beta\lambda\right) + (a-1)\log\lambda - b\lambda + \text{const.}$$

$$\therefore \log p(x_*) = \log p(x_*|\mu,\lambda) - \log p(\mu,\lambda|x_*) + \log p(\mu,\lambda).$$

$$= -\frac{1}{2}\left(\lambda(x_*-\mu)^2 - \cancel{\log\lambda}\right) + \frac{1}{2}\left(((1+\beta)\lambda(\mu-m(x_*))^2 - \log(1+\beta)\lambda\right)$$

$$- \cancel{\left(\frac{1}{2}+a-1\right)}\log\lambda + b(x_*)\lambda - \left(\frac{1}{2}+a\right)\log b(x_*)$$

$$-\frac{1}{2}\left(\beta\lambda(\mu-m)^2 - \log\beta\lambda\right) + \cancel{(a-1)\log\lambda} - b\lambda + \text{const.}$$

$$= -\frac{\lambda}{2}(x_*-\mu)^2 + \frac{\lambda}{2}(1+\beta)\left(\mu^2 - 2\mu m(x_*) + m(x_*)^2\right)$$

$$-\frac{1}{2}\left(\underbrace{\log(1+\beta)}_{\text{const.}} + \cancel{\log\lambda}\right) + b(x_*)\lambda - \left(\frac{1}{2}+a\right)\log b(x_*)$$

$$-\frac{\lambda}{2}\beta\left(\mu^2 - 2\mu m + m^2\right) + \frac{1}{2}\left(\underbrace{\log\beta}_{\text{const.}} + \cancel{\log\lambda}\right) - b\lambda + \text{const.}$$

$$= -\frac{\lambda}{2}(x_*-\mu)^2 + \frac{\lambda}{2}(1+\beta)\left(\mu^2 - 2\mu m(x_*) + m(x_*)^2\right)$$

$$+ b(x_*)\lambda - \left(\frac{1}{2}+a\right)\log b(x_*) - \frac{\lambda}{2}\beta\left(\mu^2 - 2\mu m + m^2\right) - b\lambda + \text{const.}$$

$$= -\frac{\lambda}{2}(x_*-\mu)^2 + \frac{\lambda}{2}\left(\mu^2 + \cancel{\beta\mu^2} - 2\mu(x_* + \cancel{\beta m}) + (1+\beta)m(x_*)^2\right)$$

$$+ b(x_*)\lambda - \left(\frac{1}{2}+a\right)\log b(x_*) - \frac{\lambda}{2}\beta\left(\cancel{\mu^2} - \cancel{2\mu m} + m^2\right) - b\lambda + \text{const.}$$

$$= \cancel{-\frac{\lambda}{2}(x_*-\mu)^2} + \frac{\lambda}{2}\left(\cancel{(x_*-\mu)^2} - x_*^2 + (1+\beta)m(x_*)^2\right)$$

$$+ b(x_*)\lambda - \left(\frac{1}{2}+a\right)\log b(x_*) - \frac{\lambda}{2}\beta m^2 - b\lambda + \text{const.}$$

$$= -\frac{\lambda}{2}\left(x_*^2 - (1+\beta)m(x_*)^2 + \beta m^2\right) + b(x_*)\lambda - b\lambda$$

$$-\left(\frac{1}{2}+a\right)\log b(x_*) + \text{const.}$$

$$= -\frac{\lambda}{2}\left(x_*^2 - \frac{x_*^2 + 2\beta m x_* - \beta^2 m^2}{1+\beta} + \beta m^2\right)$$

$$+ \frac{\lambda}{2}\frac{\beta}{1+\beta}\left(x_*^2 - 2mx_* + m^2\right) + \cancel{b\lambda - b\lambda}$$

$$-\left(\frac{1}{2} + a\right)\log b(x_*) + \text{const.}$$

$$= \frac{\lambda}{2(1+\beta)}\left(\cancel{-(1+\beta)x_*^2} + \cancel{x_*^2} + \cancel{2\beta m x_*} - \cancel{\beta^2 m^2} - \cancel{(1+\beta)\beta m^2}\right.$$

$$\left. + \cancel{\beta x_*^2} - \cancel{2\beta m x_*} + \cancel{\beta m^2}\right)$$

$$-\left(\frac{1}{2} + a\right)\log b(x_*) + \text{const.}$$

$$= -\left(\frac{1}{2} + a\right)\log b(x_*) + \text{const.}$$

$$= -\frac{2a+1}{2}\log\left(b\left(1 + \frac{\beta}{2b(1+\beta)}(x_* - m)^2\right)\right) + \text{const}$$

$$= -\frac{2a+1}{2}\log\left(1 + \frac{\beta}{2b(1+\beta)}(x_* - m)^2\right) + \text{const.}$$

これはStudentのt分布.

$$p(x_*) = \mathrm{St}(x_* \mid \mu_s, \lambda_s, \nu_s),$$

$$\mu_s = m, \quad \lambda_s = \frac{\beta a}{(1+\beta)b}, \quad \nu_s = 2a.$$