Predpostavimo, da je za naše podatke primeren model z normalno porazdelitvijo:

$$Y_1, Y_2, ..., Y_n|\mu, \sigma^2 \sim i.i.d.\ N(\mu, \sigma^2).$$

V nadaljevanju bomo obravnavali različice tega modela, ki se bodo razlikovale glede na predpostavke in izbiro apriornih porazdelitev.

1 Neznano upanje μ, predpostavljena znana varianca σ^2, konjugirana apriorna za μ

Če predpostavimo, da je varianca znana, se problem prevede na enoparametrično porazdelitev. Upanje μ je v tem primeru naš edini parameter, vemo pa tudi, da je v tem primeru normalna porazdelitev konjugirana apriorna porazdelitev za μ.

$$\mu|\sigma^2 \sim N(\mu_0, \sigma_0^2).$$

Zapišimo gostoti za verjetje (zaradi predpostavke iid lahko faktoriziramo)

$$p(y_1, ..., y_n|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

in izbrano apriorno porazdelitev

$$p(\mu|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}.$$

Sedaj izpeljimo aposteriorno porazdeltev za parameter μ:

$$\begin{aligned}
p(\mu|y_1, ..., y_n, \sigma^2) &= \frac{p(y_1, ..., y_n|\mu, \sigma^2)p(\mu|\sigma^2)}{p(y_1, ..., y_n)} \propto p(y_1, ..., y_n|\mu, \sigma^2)p(\mu|\sigma^2) \\
&= \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}\right] \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \\
&= (\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\frac{(\sum y_i^2 - 2\sum y_i\mu + n\mu^2)}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \\
&\propto e^{-\frac{(\sum y_i^2 - 2\sum y_i\mu + n\mu^2)}{2\sigma^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}.
\end{aligned}$$

Z malo računanja lahko izraz preoblikujemo in prepoznamo obliko normalne porazdelitve (damo na skupni imenovalec, ločimo člene z μ^2 in μ ter delimo s faktorjem poleg μ^2, da μ^2 ostane sam):

$$= e^{-\frac{(\sigma_0^2 \sum y_i^2 - \sigma_0^2 2 \sum y_i \mu + n\sigma_0^2 \mu^2) - 2\mu\mu_0\sigma^2 + \sigma^2\mu_0^2}{2\sigma^2\sigma_0^2}} \propto e^{-\frac{(n\sigma_0^2+\sigma^2)\mu^2 - 2(\sigma_0^2 \sum y_i + \mu_0\sigma^2)\mu}{2\sigma^2\sigma_0^2}} \propto e^{-\frac{\mu^2 - 2\frac{(\sigma_0^2 \sum y_i + \mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}\mu}{\frac{2\sigma^2\sigma_0^2}{(n\sigma_0^2+\sigma^2)}}}.$$

Konstantno, ki nam manjka, da bi dopolnili kvadrat v števcu ulomka, lahko prištejemo brez izgube, saj gre, ker je v eksponentu, le za množenje s konstanto, ki ne spremeni oblike. Tako lahko prepoznamo vrednosti parametrov aposteriorne porazdelitve $\mu_n = \frac{(\sigma_0^2 \sum y_i + \mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}$, $\sigma_n^2 = \frac{\sigma^2\sigma_0^2}{(n\sigma_0^2+\sigma^2)}$. Če oba delimo še z $\sigma^2\sigma_0^2$, se poenostavi v:

$$\mu|y_1, ..., y_n, \sigma^2 \sim N\left(\frac{\frac{1}{\sigma_0^2}\mu_0 + \frac{n}{\sigma^2}\overline{y}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\right)$$

Normalna porazdelitev je torej res konjugirana za normalni model z znano varianco.

Izpeljimo še aposteriorno napovedno porazdelitev (μ_n in σ_n^2 sta parametra aposteriorne, ki smo jo izpeljavi):

$$p(y_{new}|y, \sigma^2) = \int_{-\infty}^{\infty} p(y_{new}|\mu, \sigma^2) p(\mu|y, \sigma^2) d\mu \propto \int e^{-\frac{(y_{new}-\mu)^2}{2\sigma^2}} e^{-\frac{(\mu-\mu_n)^2}{2\sigma_n^2}} d\mu$$

$$= \int e^{-\frac{\sigma_n^2 y_{new}^2 - 2\sigma_n^2 y_{new}\mu + \sigma_n^2\mu^2 + \sigma^2\mu^2 - 2\sigma^2\mu\mu_0 + \sigma^2\mu_0^2}{2\sigma^2\sigma_n^2}} d\mu$$

Faktorje, ki ne vsebujejo μ vzamemo ven iz integrala:

$$= e^{-\frac{\sigma_n^2 y_{new}^2 + \sigma^2\mu_0^2}{2\sigma^2\sigma_n^2}} \int_{\mathbb{R}} e^{-\frac{\mu^2 - 2(\sigma^2\mu_0 + \sigma_n^2 y_{new})/(\sigma_n^2+\sigma^2)\mu}{2\sigma^2\sigma_n^2/(\sigma_n^2+\sigma^2)}} d\mu$$

Če izraz množimo z $e^{-\frac{((\sigma^2\mu_0+\sigma_n^2 y_{new})/(\sigma_n^2+\sigma^2))^2}{2\sigma^2\sigma_n^2/(\sigma_n^2+\sigma^2)}}$ in njegovo obratno vrednostjo (da vrednost ostane nespremenjena), bomo s prvim izrazom dopolnili kvadrat znotraj eksponenta. Če ta integral pomnožimo še z ustrezno konstanto, ki je neodvisna od y_{new} in μ ter tako ne spremeni proporcionalnosti, dobimo integral gostote normalne porazdelitve, za kateraga vemo, da mora biti 1. Ostane nam:

$$\propto e^{-\frac{\sigma_n^2 y_{new}^2 + \sigma^2\mu_0^2}{2\sigma^2\sigma_n^2}} e^{\frac{((\sigma^2\mu_0+\sigma_n^2 y_{new})/(\sigma_n^2+\sigma^2))^2}{2\sigma^2\sigma_n^2/(\sigma_n^2+\sigma^2)}}$$

in, z malo dodatnega računanja

$$= e^{-\frac{(y_{new}-\mu_n)^2}{2(\sigma^2+\sigma_n^2)}}.$$

Aposteriorna napovedna porazdelitev je torej: $y_{new}|y, \sigma^2 \sim N(\mu_n, \sigma^2 + \sigma_n^2)$.

2 Neznana varianca σ^2, predpostavljeno znano upanje μ, konjugirana apriorna za σ^2

V tem primeru je σ^2 naš edini parameter, ki ga bomo označili kar z $\theta = \sigma^2$. Vemo pa tudi, konjugirana apriorna porazdelitev porazdelitev inverzna Gama

$$\theta|\mu \sim \mathcal{IG}(\alpha_0, \beta_0)$$

njena gostota pa

$$p(\theta|\mu) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\theta^{-\alpha_0-1}e^{-\frac{\beta}{\theta}}.$$

Sedaj izpeljimo aposteriorno porazdeltev za parameter θ:

$$\begin{aligned}
p(\theta|y_1, ..., y_n, \mu) &= \frac{p(y_1, ..., y_n|\mu, \theta)p(\theta|\mu)}{p(y_1, ..., y_n)} \propto p(y_1, ..., y_n|\mu, \theta)p(\theta|\mu) \\
&= \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(y_i-\mu)^2}{2\theta}}\right] \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\theta^{-\alpha_0-1}e^{-\frac{\beta}{\theta}} \\
&\propto (\frac{1}{\sqrt{2\pi\theta}})^n e^{-\frac{\sum(y_i-\mu)^2}{2\theta}} \theta^{-\alpha_0-1}e^{-\frac{\beta}{\theta}} \\
&\propto \theta^{-\frac{n}{2}-\alpha_0-1}e^{-\frac{\frac{1}{2}\sum(y_i-\mu)^2+\beta_0}{\theta}}.
\end{aligned}$$

Ko prepoznamo obliko porazdelitve inverzna gama in namesto θ spet pišemo σ^2, sledi:

$$\sigma^2|y_1, ..., y_n, \mu \sim \mathcal{IG}\left(\frac{n}{2} + \alpha_0, \frac{1}{2}\sum(y_i - \mu)^2 + \beta_0\right).$$

3 Neznano upanje μ, neznana varianca σ^2, odvisni apriorni za μ in σ^2

Če pri normalnem modelu modeliramo oba parametra, obstaja samo en način, kako določiti apriorne porazdelitve, če želimo obvladljivo (analitično) aposteriorno gostoto:

$\mu|\sigma^2 \sim N(\mu_0, \frac{\sigma^2}{\kappa_0})$ in

$\sigma^2 \sim \mathcal{IG}(\alpha_0, \beta_0)$.

Ta omejitev je sicer smiselna, saj mora biti pri predpostavki normalnega modela in znani varianci σ^2 negotovost našega mnenja o μ za nek faktor manjša od variance populacije. κ_0 torej predstavlja navidezno število vzorcev, na katerih temelji naše apriorno mnenje o μ.

To je naš prvi primer večparametričnega modela in s tem več oz. multivariatne aposteriorne porazdelitve. V praksi se z multivariatnimi porazdelitvami redko spopademo neposredno, razen, ko so te analitično obvladljive (npr. multivariatna normalna porazdelitev). V večini primerov porazdelitev faktoriziramo na bolj obvladljive faktorje:

$$p(\mu, \sigma^2|y) = p(\mu|y, \sigma^2)p(\sigma^2|y).$$

Izračun prvega faktorja je relativno preprost - gre za primer modela z znano varianco, ki smo ga že izpeljali, le vstavimo $\sigma_0^2 = \frac{\sigma^2}{\kappa_0}$.

Drugi faktor pa bomo izračunali preko marginalizacije skupne porazdelitve $p(\mu, \sigma^2|y) = \frac{p(\mu,\sigma^2,y)}{p(y)} \propto p(\mu, \sigma^2, y) = p(\sigma^2)p(\mu|\sigma^2)p(y|\mu, \sigma^2)$ oziroma

$$p(\sigma^2|y) \propto \int p(\sigma^2)p(\mu|\sigma^2)p(y|\mu, \sigma^2)d\mu.$$

Spet bomo pisali θ namesto σ^2, da poenostavimo. Osredotočimo se na izraz znotraj zgornjega integrala in že na začetku odmislimo vse faktorje, ki niso odvisni od σ^2:

$$p(\sigma^2)p(\mu|\sigma^2)p(y|\mu, \sigma^2) \propto \overbrace{\theta^{-(\alpha_0+1)}\exp(-\beta_0/\theta)}^{\text{apriorna za } \sigma^2} \overbrace{\frac{1}{\sqrt{\theta}}\exp(-\kappa_0(\mu-\mu_0)^2/(2\theta))}^{\text{pogojna apriorna } \mu|\theta} \overbrace{\theta^{-n/2}\exp(-\sum(y_i-\mu)^2/(2\theta))}^{\text{model}}$$

Sedaj uporabimo

$\sum(y_i-\mu)^2 = \sum(y_i-\overline{y}+\overline{y}-\mu)^2 = \sum(y_i-\overline{y})^2 + 2\sum(y_i-\overline{y})(\overline{y}-\mu) + \sum(\overline{y}-\mu)^2 = \sum(y_i-\overline{y})^2 + n(\overline{y}-\mu)^2$,

da dobimo

$= \theta^{-(\alpha_0+1)}\exp(-\beta_0/\theta)\frac{1}{\sqrt{\theta}}\theta^{-n/2}\exp(-\sum(y_i-\overline{y})^2/2\theta)\exp(-\kappa_0(\mu-\mu_0)^2/2\theta)\exp(-n(\overline{y}-\mu)^2/2\theta)$

$= \theta^{-(\alpha_0+1)}\exp(-\beta_0/\theta)\frac{1}{\sqrt{\theta}}\theta^{n/2}\exp(-\sum(y_i-\overline{y})^2/2\theta)\exp(-(\kappa_0(\mu-\mu_0)^2 + n(\overline{y}-\mu)^2)/2\theta)$,

pri čemer lahko zadnji faktor, ki je edini odvisen od μ, preuredimo

$= \exp(-(\kappa_0(\mu-\mu_0)^2 + n(\overline{y}-\mu)^2)/2\theta)$

$= \exp(-(\kappa_0\mu^2 - 2\kappa_0\mu\mu_0 + \kappa_0\mu_0^2 + n\overline{y}^2 - 2n\overline{y}\mu + n\mu^2)/2\theta)$

$= \exp(-((\kappa_0+n)\mu^2 - 2(\kappa_0\mu_0 + n\overline{y})\mu + \kappa_0\mu_0^2 + n\overline{y}^2)/2\theta)$

$= \exp(-(\mu^2 - 2\frac{\kappa_0\mu_0+n\overline{y}}{(\kappa_0+n)}\mu + \frac{\kappa_0\mu_0^2+n\overline{y}^2}{(\kappa_0+n)})/\frac{2\theta}{(\kappa_0+n)})$.

Da bi v stevcu dobili popoln kvadrat, potrebujemo $(\frac{\kappa_0\mu_0+n\overline{y}}{(\kappa_0+n)})^2 = \frac{\kappa_0^2\mu_0^2+\kappa_0\mu_0 n\overline{y}+n^2\overline{y}^2}{(\kappa_0+n)^2}$, pri čemer imamo trenutno $\frac{(\kappa_0+n)(\kappa_0\mu_0^2+n\overline{y}^2)}{(\kappa_0+n)^2} = \frac{(\kappa_0^2\mu_0^2+\kappa_0 n\overline{y}^2+n\kappa_0\mu_0^2+n^2\overline{y}^2)}{(\kappa_0+n)^2}$. Manjka nam torej $\frac{(\kappa_0\mu_0 n\overline{y}-\kappa_0 n\overline{y}^2-n\kappa_0\mu_0^2)}{(\kappa_0+n)^2}$, da pa bi to spravili v enačbo, moramo enako tudi odšteti, zato vse skupaj pomnozimo s členom $\exp(-\frac{(\kappa_0\mu_0 n\overline{y}-\kappa_0 n\overline{y}^2-n\kappa_0\mu_0^2)}{(\kappa_0+n)2\theta}) = \exp(-\frac{\kappa_0 n(\mu_0-\overline{y})^2}{2(\kappa_0+n)}/\theta)$.

Da dobimo ustrezno obliko normalne porazdelitve (ki se integrira v konstanto) potrebujemo še s $c\frac{1}{\sqrt{\theta}}$ in ustrezen popravek z množenjem z $\sqrt{\theta}$.

Faktor z μ smo torej preoblikovali v gostoto normalne, zato se pri marginalizaciji preko μ izniči v 1. Iz konstantnih faktorjev (z vidika μ), ki ostanejo, lahko razberemo marginalno aposteriorno variance:

$\propto \theta^{-(\alpha_0+1)} \exp(-\beta_0/\theta)\frac{1}{\sqrt{\theta}}\theta^{n/2}\exp(-\sum(y_i-\overline{y})^2/2\theta)\sqrt{\theta}\exp(-\frac{\kappa_0 n(\mu_0-\overline{y})^2}{2(\kappa_0+n)}/\theta)$

$= \theta^{-(\alpha_0+\frac{n}{2}+1)}\exp(-\frac{\beta_0+\frac{1}{2}\sum(y_i-\overline{y})^2+\frac{1}{2}\frac{\kappa_0 n(\mu_0-\overline{y})^2}{(\kappa_0+n)}}{\theta})$

Torej,

$$\begin{aligned}
\mu|y,\sigma^2 &\sim N(\frac{\kappa_0\mu_0+n\overline{y}}{\kappa_0+n}, \frac{\sigma^2}{\kappa_0+n}) \\
\sigma^2|y &\sim \mathcal{IG}\left(\alpha_0+\frac{n}{2}, \beta_0+\frac{1}{2}(\sum(y_i-\overline{y})^2+\frac{\kappa_0 n}{\kappa_0+n}(\overline{y}-\mu_0)^2)\right).
\end{aligned}$$

Posebnega pomena je tudi marginalna aposteriorna $\mu|y$, ki pa jo bomo izpeljali le za poseben primer $p(\mu|\sigma^2) \propto 1$, $p(\sigma^2) \propto \frac{1}{\sigma^2}$ (spet bomo bisali θ namesto σ^2).

$p(\mu,\sigma^2|y) \propto \frac{1}{\sigma^2}\theta^{-n/2}\exp(-\sum(y_i-\mu)^2/(2\theta))$

$= \theta^{-(1+\frac{n}{2})}\exp(-\sum(y_i-\overline{y})^2/2\theta)\exp(-n(\overline{y}-\mu)^2/2\theta)$

$= \theta^{-(1+\frac{n}{2})}\exp(-\frac{\sum(y_i-\overline{y})^2+n(\overline{y}-\mu)^2}{2\theta})$

$= \theta^A\exp(\frac{B}{\theta})$, kjer je $A = -(1+\frac{n}{2})$ in $B = (-\frac{\sum(y_i-\overline{y})^2+n(\overline{y}-\mu)^2}{2})$.

Podporna izpeljava

$$\int_0^\infty \theta^A e^{\frac{B}{\theta}} d\theta$$

S substitucijo $x = \frac{B}{\theta}$, $dx = -\frac{B}{\theta^2}d\theta$ pridemo do

$$\int_0^\infty \theta^A e^x - \frac{\theta^2}{B} dx \propto \frac{1}{B} \int_0^\infty \theta^{A+2} e^x dx = \frac{1}{B} \int_0^\infty (\frac{B}{x})^{A+2} e^x dx = B^{A+1} \int_0^\infty x^{-A-2} e^x dx$$

$$\propto B^{A+1} \propto (\sum (y_i - \overline{y})^2 + n(\overline{y} - \mu)^2)^{-\frac{n}{2}} = (1 + \frac{n(\overline{y} - \mu)^2}{vs^2})^{-\frac{v+1}{2}},$$

kjer sta $s^2 = \frac{1}{n-1} \sum (y_i - \overline{y})^2$ in $v = n - 1$.

Če vpeljemo $t = \frac{\mu - \overline{y}}{s/\sqrt{n}}$, dobimo

$$(1 + \frac{t^2}{v})^{-\frac{v+1}{2}}$$

Torej, marginalna aposteriorna porazdelitev μ je skalirana t-porazdelitev z $n - 1$ prostostnimi stopnjami.