Prove Gamma-Poisson conjugation

likelihood 令 $x \sim \text{Poisson}(\lambda)$ pmf: $p(x|\lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

prior $\lambda \sim \text{Gamma}(\alpha, \beta)$ $p(\lambda) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$

$$p(\lambda|x) \propto p(x|\lambda) \cdot p(\lambda)$$

$$\propto \frac{\lambda^x e^{-\lambda}}{x!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{(\alpha+x)-1} e^{-(\beta+1)\lambda}$$

$$\text{Gamma}(\lambda|\alpha', \beta') = \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \lambda^{\alpha'-1} e^{-\beta'\lambda} \quad \alpha' = \alpha + x \quad \beta' = \beta + 1$$

∴ Gamma 分布是 Poisson 分布的共軛先驗數

因爲 posterior 仍是 Gamma 分布

---

prove mean and variance with Gaussian prior $w \sim N(\mu_0, \Lambda_0^{-1})$

prior $p(w) = N(\mu_0, \Lambda_0^{-1})$

$$p(w) \propto \exp\left(-\frac{1}{2}(w-\mu_0)^T \Lambda_0 (w-\mu_0)\right)$$

$$= \exp\left(-\frac{1}{2}\left[w^T \Lambda_0 w - 2\mu_0^T \Lambda_0 w + \mu_0^T \Lambda_0 \mu_0\right]\right)$$

$$(\phi w)^T(\phi w) = w^T \phi^T \phi w$$

likelihood $p(y|w) = N(\phi w, aI)$

$$p(y|w) \propto \exp\left(\frac{-1}{2a}(y-\phi w)^T(y-\phi w)\right)$$

$$= \exp\left(-\frac{1}{2a}\left[y^T y - 2y^T \Phi w + w^T \phi^T \phi w\right]\right)$$

Bayes theorem $p(w|y) \propto p(y|w) \cdot p(w)$

$$p(w|y) = \exp\left(-\frac{1}{2}\left[w^T \Lambda_0 w - 2\mu_0^T \Lambda_0 w + \mu_0^T \Lambda_0 \mu_0\right]\right) \times$$
$$\exp\left(-\frac{1}{2a}\left[y^T y - 2y^T \Phi w + w^T \phi^T \phi w\right]\right)$$

$$\mu_0^T \Lambda_0 w + \frac{1}{a} y^T \Phi w$$

$$\log p(w|y) = -\frac{1}{2}\left[w^T \Lambda_0 w - 2\mu_0^T \Lambda_0 w + \mu_0^T \Lambda_0 \mu_0\right] +$$
$$-\frac{1}{2a}\left[y^T y - 2y^T \Phi w + w^T \phi^T \phi w\right]$$

$$w^T \mu_0^T \Lambda_0 + w^T \frac{1}{a} y^T \phi$$
$$= w^T\left(\mu_0 \Lambda_0 + \frac{1}{a} \phi^T y\right)$$

$$= -\frac{1}{2} w^T \left(\Lambda_0 + \frac{1}{a}\phi^T\phi\right) w + w^T\left(\frac{1}{a}\phi^T y + \Lambda_0 \mu_0\right) + \text{costant}$$

$$\log p(w) = -\frac{1}{2} w^T \Lambda w + w^T b \qquad \Lambda = \Sigma_N^{-1}$$
$$\mu_N = \Lambda^{-1} \cdot b$$

$$\Sigma_N = \left(\Lambda_0 + \frac{1}{a}\phi^T\phi\right)^{-1}$$
$$\mu_N = \Sigma_N\left(\frac{1}{a}\phi^T y + \Lambda_0 \mu_0\right)$$

$\boxed{\text{Bayes Linear Regression}}$

precision 初始值

initial prior $w \sim N(0, b^{-1} I) \Rightarrow \mu = 0,\ \Lambda_0 = b \cdot I$

$$\Sigma_N = \left(bI + \frac{1}{a}\phi^T\phi\right)^{-1}$$

每筆資料帶來的資訊

$$\mu = \Sigma_N\left(\frac{1}{a}\phi^T y\right)$$

ϕ column vector $(n \times 1)$

每筆資料對 mean 的影響 $m = \frac{1}{a}\sum y_i \phi(x_i)$