Soln 1

i) Now, Beta distribution is given as:

$$B(u \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \, u^{a-1} (1-u)^{b-1}$$

and $E[u] = \dfrac{a}{a+b}$

Now, Variance can be written as

$$Var[u] = E[u^2] - \left(E[u]\right)^2$$

we have $E[u]$, so let's find $E[u^2]$

$$\therefore \quad E[u^2] = \int_0^1 u^2 \, B(u \mid a, b) \, du$$

$$= \int_0^1 u^2 \, \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \, u^{a-1}(1-u)^{b-1} \, du$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \int_0^1 u^{(a+2)-1} (1-u)^{b-1} \, dx$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \underbrace{\int_0^1 \frac{\Gamma(a+b+2)}{\Gamma(a+2)\Gamma(b)} \cdot u^{(a+2-1)} (1-u)^{b-1} \, dx}_{(\text{will integrate to } 1)}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \, \frac{\Gamma(a+2)\,\Gamma(b)}{\Gamma(a+b+2)}$$

1

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{(a+1)(a)}{(a+b+1)(a+b)} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\boxed{E[u^2] = \frac{(a+1)(a)}{(a+b+1)(a+b)}}$$

$$\therefore \quad Var[u] = E[u^2] - (E[u])^2$$

$$= \frac{(a+1)(a)}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2}$$

$$= \frac{a}{a+b}\left[\frac{a+1}{a+b+1} - \frac{a}{a+b}\right]$$

$$= \frac{a}{a+b}\left[\frac{a^2+a+ab+b - a^2 - ab - a}{(a+b+1)(a+b)}\right]$$

$$= \frac{a}{a+b}\left[\frac{b}{(a+b)(a+b+1)}\right]$$

$$= \frac{ab}{(a+b)^2(a+b+1)}$$

$$\therefore \quad \boxed{Var[u] = \frac{ab}{(a+b)^2(a+b+1)}}$$

ii) Dirichlet distribution is given as:

$$D(u|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_K)} \prod_{R=1}^{K} u_R^{\alpha_R - 1}$$

where $u = (u_1, u_2, \dots u_j \dots u_K)$

Now, $E[u_j] = \int_0^1 u_j \, D(u|\alpha) \, du$

$$= \int_0^1 u_j \, \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_K)} \prod_{R=1}^{K} u_R^{\alpha_R - 1} \, du$$

$$= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_K)} \int_0^1 u_j^{\alpha_j} \prod_{\substack{K=1 \\ \text{and } K \neq j}}^{K} u_R^{\alpha_R - 1} \, du$$

Now, suppose $\omega_j = \alpha_j + 1$ and $\forall_{K \neq j} \omega_K = \alpha_K$

$\therefore \quad \omega_0 = \sum_{R=1}^{K} \omega_K = \sum_{R=1}^{K} \alpha_K + 1 = \alpha_0 + 1$

$\therefore \quad \Gamma(\omega_0) = \alpha_0 \Gamma(\alpha_0)$ and $\Gamma(\omega_j) = \alpha_j \Gamma(\alpha_j)$

$\therefore \quad E[u_j] = \frac{\alpha_j}{\alpha_0} \frac{\Gamma(\omega_0)}{\Gamma(\omega_1)\Gamma(\omega_2)\dots\Gamma(\omega_K)} \int_0^1 \prod_{R=1}^{K} u_R^{\omega_R - 1} \cdot du$

$$\therefore E[u_j] = \frac{x_j}{x_0} \int_0^1 \underbrace{\frac{\Gamma(w_0)}{\Gamma(w_1) \cdots \Gamma(w_K)} \prod_{R=1}^{K} u_R^{w_R - 1}}_{} du$$

integrates to 1 because it
is a dirichlet distribution.

$$\therefore \boxed{E[u_j] = \frac{x_j}{x_0}}$$

Soln 2

1) From the text, dirichlet distribution is given as:

$$Dih(u|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{i=1}^{K} u_K^{\alpha_K - 1} \qquad - ①$$

also the Posterior is given as:

$$p(u|data) = \frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_1 + m_1)\cdots\Gamma(\alpha_K + m_K)} \prod_{K=1}^{K} u_R^{\alpha_R + m_R - 1} \qquad - ②$$

Now, the predictive distribution is given as:

$$P(\text{next word is } w_j \mid data) = \int_u P(u|data) \cdot P(x_{n+1} = w_j \mid u) \cdot du$$

$$= \int_u u_j \, P(u|data)\, du$$

$$= E[u_j \mid data] \quad (\text{This is expected value of } u_j \text{ from posterior distribution})$$

From part ii) of soln 1, we get that:

$$E[u_j] = \frac{\alpha_j'}{\alpha_0'}$$

From ②, we have $\alpha_j' = \alpha_j + m_j$, $\alpha_0' = \alpha_0 + N$

$$\therefore \boxed{P(x_{N+1} = w_j \mid data) = \frac{\alpha_j + m_j}{\alpha_0 + N}}$$

5

ii) $P(Data | \alpha) = \int_{\mu} P(data | \mu) \, P(\mu | \alpha) \cdot d\mu$

$= \int_{\mu} \prod_{R=1}^{K} \mu_R^{m_R} \cdot \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \cdot \Gamma(\alpha_R)} \prod_{R=1}^{K} \mu_R^{\alpha_R - 1} \cdot d\mu$

$= \frac{\Gamma(\alpha_0)}{\prod_{K=1}^{K} \Gamma(\alpha_R)} \int_{\mu} \prod_{R=1}^{K} \mu_R^{\alpha_R + m_R - 1} \cdot d\mu$

$= \frac{\Gamma(\alpha_0)}{\prod_{R=1}^{K} \Gamma(\alpha_R)} \cdot \frac{\prod_{R=1}^{K} \Gamma(\alpha_K + m_R)}{\Gamma(\alpha_0 + N)} \underbrace{\int_{\mu} \frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_1 + m_1) \cdot \Gamma(\alpha_R + m_R)} \cdot \prod \mu_R^{\alpha_R + m_R - 1} \, d\mu}$

Normalized Dirichlet distribution hence this integral will be 1.

$\therefore \boxed{P(data | \alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + N)} \cdot \frac{\prod_{K=1}^{K} \Gamma(\alpha_R + m_R)}{\prod_{R=1}^{K} \Gamma(\alpha_R)}}$

6

Soln 3

Normal distribution can be written as:

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Let, $\frac{1}{\sigma^2} = \beta$

$\therefore \quad p(x) = \frac{\sqrt{\beta}}{\sqrt{2\pi}} \, e^{-\frac{\beta}{2}(x-\mu)^2}$

$\therefore \quad p(x) = \frac{\sqrt{\beta}}{\sqrt{2\pi}} \, e^{-\frac{\beta}{2}x^2} \cdot e^{\beta\mu x} \cdot e^{-\frac{\beta}{2}\mu^2} \quad - \text{①}$

Liklihood function can be written as

$$L = p(data \mid model) = C \, e^{-\frac{\beta}{2}\Sigma x_i^2} \cdot e^{\beta\mu\Sigma x_i} \cdot e^{-\frac{N\beta}{2}\mu^2}$$

Now, Prior is also a normal distribution:

$\therefore \quad p(model) = \frac{\sqrt{\beta_0}}{\sqrt{2\pi}} \, e^{-\frac{\beta_0}{2}(\mu-\mu_0)^2} \qquad ; \quad \beta_0 = \frac{1}{\sigma_0^2}$

$\therefore \quad p(model) = \frac{\sqrt{\beta_0}}{\sqrt{2\pi}} \, e^{-\frac{\beta_0}{2}\mu^2} \cdot e^{\beta_0\mu\mu_0} \cdot e^{-\frac{\beta_0}{2}\mu_0^2}$

Posterior $\propto$ Liklihood $\times$ Prior.

$\therefore \quad p(model \mid data) \propto p(data \mid model) \times p(model)$

$\therefore\ p(\text{model} \mid \text{data}) \propto e^{-\frac{\beta}{2}\Sigma x_i^2} \cdot e^{\beta\mu\Sigma x_i} \cdot e^{-\frac{N}{2}\beta\mu^2} \cdot$

$$e^{-\frac{\beta_0}{2}\mu^2} \cdot e^{\beta_0\mu\mu_0} \cdot e^{-\frac{\beta_0}{2}\mu_0^2}$$

$\therefore\ p(\text{model} \mid \text{data}) \propto e^{-\mu^2\left(\frac{\beta_0 + N\beta}{2}\right)} \cdot e^{\mu[\beta_0\mu_0 + \beta\Sigma x_i]}$

$$\cdot\ e^{-\frac{1}{2}\left[\beta_0\mu_0^2 + \beta\Sigma x_i^2\right]}$$

Now, comparing this to ①, it resembles a normal distribution. Hence, $\mu \sim N(\mu_0, \sigma_0^2)$ is conjugate prior to the likelihood function.

$\therefore$ let, $p(\text{model} \mid \text{data}) = \dfrac{\sqrt{\beta'}}{\sqrt{2\pi}} \cdot e^{-\frac{\beta'}{2}\mu^2} \cdot e^{\beta'\mu\mu'} \cdot e^{\frac{\beta'}{2}\mu'^2}$

Solving for μ' and β',

$$\dfrac{\beta'}{2} = \dfrac{\beta_0 + N\beta}{2} \quad \therefore \boxed{\beta' = \beta_0 + N\beta}$$

and $\beta'\mu' = \beta_0\mu_0 + \beta\Sigma x_i \quad \therefore \boxed{\mu' = \dfrac{\beta_0\mu_0 + \beta\Sigma x_i}{\beta_0 + N\beta}}$

$\therefore$ posterior $= p(\text{model} \mid \text{data})$

$$= \dfrac{\sqrt{\beta_0 + N\beta}}{\sqrt{2\pi}}\ e^{-\frac{(\beta_0 + N\beta)}{2}\left[\mu - \frac{\beta_0 + \mu_0 + \beta\Sigma x_i}{\beta_0 + N\beta}\right]^2}$$

8

Soln 4

i) Poisson distribution is given as:

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

data : $x_1, x_2, \ldots x_N$

∴ Liklihood function is given as:

$$L = p(data|\lambda) = \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

∴ $\log L = \log\left(\prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)$

$$= \sum_{i=1}^{N} \log\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)$$

$$= \sum_{i=1}^{N} \left[\log e^{-\lambda} + x_i \log\lambda - \log(x_i!)\right]$$

$$= -N\lambda + \log\lambda \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} \log(x_i!)$$

Now, for estimating λ we need to do $\frac{d(\log L)}{d\lambda} = 0$

∴ $-N + \frac{1}{\lambda}\sum_{i=1}^{N} x_i = 0$

∴ $\boxed{\hat{\lambda}_{ML} = \frac{\sum_{i=1}^{N} x_i}{N}}$

9