Soln 1

i) Now, Beta distribution is given as:

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

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

Now, Variance can be written as

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

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

$$\therefore E[\mu^2] = \int_0^1 \mu^2 \, \beta(\mu \mid a,b)\, d\mu$$

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

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

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

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

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

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

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

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

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

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

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

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

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

ii) Dirichlet distribution is given as:

$$D(\mu \mid \alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$

where $\mu = (\mu_1, \mu_2, \dots \mu_j \dots \mu_K)$

Now, $E[\mu_j] = \int_0^1 \mu_j \, D(\mu \mid \alpha) \, d\mu$

$$= \int_0^1 \mu_j \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \, d\mu$$

$$= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)} \int_0^1 \mu_j^{\alpha_j} \prod_{\substack{k=1 \\ \text{and } k \neq j}}^{K} \mu_k^{\alpha_k - 1} \, d\mu$$

Now, suppose $\omega_j = \alpha_j + 1$ and $\forall_{k \neq j} \; \omega_k = \alpha_k$

$\therefore \quad \omega_0 = \sum_{k=1}^{K} \omega_k = \sum_{k=1}^{K} \alpha_k + 1 = \alpha_0 + 1$

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

$$\therefore \quad E[\mu_j] = \frac{\alpha_j}{\alpha_0} \frac{\Gamma(\omega_0)}{\Gamma(\omega_1)\Gamma(\omega_2) \dots \Gamma(\omega_K)} \int_0^1 \prod_{k=1}^{K} \mu_k^{\omega_k - 1} \, d\mu$$

$$\therefore \quad E[\mu_j] = \frac{\alpha_j}{\alpha_0} \underbrace{\int_0^1 \frac{\Gamma(\omega_0)}{\Gamma(\omega_1) \quad \Gamma(\omega_K)} \prod_{k=1}^{K} \mu_k^{\omega_k - 1} \, d\mu}_{\text{integrates to 1 because it is a dirichlet distribution.}}$$

$$\therefore \quad \boxed{E[\mu_j] = \frac{\alpha_j}{\alpha_0}}$$

Soln 2

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

$$Dir(u|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{i=1}^{K} u_k^{\alpha_k - 1} \quad - (1)$$

also the Posterior is given as:

$$p(u|\text{data}) = \frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_1 + m_1)\cdots\Gamma(\alpha_K + m_K)} \prod_{k=1}^{K} u_k^{\alpha_k + m_k - 1} \quad - (2)$$

now, the predictive distribution is given as:

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

$$= \int_u u_j \; P(u|\text{data})\, du$$

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

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

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

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

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

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

$$= \int_{\mu} \prod_{k=1}^{K} \mu_k^{m_k} \cdot \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \cdot d\mu$$

$$= \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \int_{\mu} \prod_{k=1}^{K} \mu_k^{\alpha_k + m_k - 1} \cdot d\mu$$

$$= \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{K} (\alpha_k)} \cdot \frac{\prod_{k=1}^{K} \Gamma(\alpha_k + m_k)}{\Gamma(\alpha_0 + N)} \int_{\mu} \underbrace{\frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_1 + m_1) \cdots \Gamma(\alpha_k + m_k)} \cdot \prod \mu_k^{\alpha_k + m_k - 1}\, d\mu}$$

Normalized Dirichlet distribution
hence this integral
will be 1.

$$\therefore \boxed{P(\text{data} \mid \alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + N)} \cdot \frac{\prod_{k=1}^{K} \Gamma(\alpha_k + m_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)}}$$

Soln 3

Normal distribution can be written as:

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

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

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

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

Liklihood function can be written as

$$L = p(\text{data} \mid \text{model}) = c\, e^{-\beta/2 \sum x_i^2} \cdot e^{\beta\mu \sum x_i} \cdot e^{-\frac{N\beta}{2}\mu^2}$$

Now, Prior is also a normal distribution:

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

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

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

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

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

$$\therefore \quad p(\text{model} \mid \text{data}) \propto e^{-\mu^2\left(\frac{\beta_0 + N\beta}{2}\right)} \cdot e^{\mu[\beta_0\mu_0 + \beta \Sigma x_i]} \cdot e^{-\frac{1}{2}[\beta_0 \mu_0^2 + \beta \Sigma x_i^2]}$$

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

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

Solving for μ' and β',

$$\frac{\beta'}{2} = \frac{\beta_0 + N\beta}{2} \quad \Bigg| \quad \therefore \boxed{\beta' = \beta_0 + N\beta}$$

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

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

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

Soln 4

i) Poisson distribution is given as:

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

data: $x_1, x_2, \ldots x_N$

$\therefore$ Liklihood function is given as:

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

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

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

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

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

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

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

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

ii) From part i) we obtained,

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

$$\therefore \ E[\hat{\lambda}_{ML}] = E\left[\frac{1}{N} \sum_{i=1}^{N} x_i \right]$$

$$= \frac{1}{N} E\left[\sum_{i=1}^{N} x_i \right]$$

$$= \frac{1}{N} . N . E[x] \quad (\text{given } E[x] = \lambda)$$

$$\boxed{E[\hat{\lambda}_{ML}] = \lambda}$$

now, $$Var[\hat{\lambda}_{ML}] = Var\left[\frac{1}{N} \sum_{i=1}^{N} x_i \right]$$

$$= \frac{1}{N^2} Var\left[\sum_{i=1}^{N} x_i \right]$$

$$= \frac{1}{N^2} N \ Var[x] \quad (\text{given } Var[x] = \lambda)$$

$$\boxed{Var(\hat{\lambda}_{ML}) = \frac{\lambda}{N}}$$

Soln 5

i) To prove that gamma is conjugate to poisson, we need to prove that given a gamma prior, we obtain a gamma functional form of posterior as well.

Now, posterior can be written as:

Posterior $\propto$ Liklihood $\times$ Prior

$\therefore \; p(\lambda \mid data) \propto p(data \mid \lambda) \times p(\lambda)$

now, from soln 4 we got:

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

$$= \frac{e^{-N\lambda} \cdot \lambda^{x_1 + x_2 \dots + x_N}}{x_1! \, x_2! \dots x_N!}$$

$$= K \, e^{-N\lambda} \cdot \lambda^{\Sigma x_i}$$

(where K is constant)

and we have,

$$p(\lambda) = \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda^{a_0 - 1} e^{-b_0 \lambda} \quad \text{(given)}$$

$$\therefore \; p(\lambda \mid data) \propto K \, e^{-N\lambda} \lambda^{\Sigma x_i} \cdot \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda^{a_0 - 1} e^{-b_0 \lambda}$$

$\therefore \quad p(\lambda | data) \propto e^{-\lambda(N+b_0)} \cdot \lambda^{(\Sigma x_i + a_0)-1}$

$\because$ Now, $p(\lambda | data)$ has same functional form as that of its prior hence we can say that our prior is conjugate to likelihood i.e gamma distribution is prior to poisson distribution.

$\therefore \quad p(\lambda | data) = c \; e^{-\lambda(N+b_0)} \cdot \lambda^{(\Sigma x_i + a_0)-1}$

$\therefore$ Comparing with gamma distribution to compute c,

$$a_N = \sum_{i=1}^{N} x_i + a_0$$

$$b_N = N + b_0$$

$$\therefore \quad \boxed{p(\lambda | data) = \frac{(N+b_0)^{(\Sigma x_i + a_0)}}{\Gamma(\Sigma x_i + a_0)} e^{-\lambda(N+b_0)} \lambda^{(\Sigma x_i + a_0)-1}}$$

ii) From result of i) let,

$$\frac{(N+b_0)^{(\Sigma x_i + a_0)}}{\Gamma(\Sigma x_i + a_0)} = C \quad \text{(constant)}$$

and, let $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ (mean)

$\therefore \Sigma x_i = N\bar{x}$

Substituting these we get,

$$p(\lambda \mid \text{data}) = C\, e^{-\lambda(N+b_0)} \lambda^{(N\bar{x}+a_0-1)}$$

To get MAP estimate, we need to do

$$\frac{d}{d\lambda} p(\lambda \mid \text{data}) = 0$$

$$\therefore C\left[-(N+b_0)e^{-\lambda(N+b_0)}\lambda^{(N\bar{x}+a_0-1)} + (N\bar{x}+a_0-1)e^{-\lambda(N+b_0)}\lambda^{(N\bar{x}+a_0-2)}\right]$$

$$\therefore \left[-(N+b_0) + \frac{(N\bar{x}+a_0-1)}{\lambda}\right] = 0$$

$$\therefore \boxed{\hat{\lambda}_{MAP} = \frac{N\bar{x}+a_0-1}{N+b_0}}$$