Connjugate Prior

Why are they useful and proofs of thier righteousness

Ha Vu

07/26/2023

Contents

1 Why conjugate prior is nice?

1.1 What is conjugate prior in the first place?

In Bayesian inference, X represents the observed variable, which has a distribution that is parameterized by θ, i.e. $X \sim f_\theta(x)$. Bayesian inference assumes that $\theta \sim g(\theta)$, which is called a prior distribution. In most of the cases for application, the goal is to infer the values of θ that maximize the posterior probability of θ, i.e. $\arg\max_\theta P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int P(X|\hat{\theta})\, d\hat{\theta}}$ Conjugate prior is the case where the prior distribution of $P(\theta)$ is chosen so that the posterior distribution $P(\theta|X)$ is of the same distribution as the prior.

1.2 Why is it nice?

Conjugate is nice because we KNOW analytical form of the posterior distribution.

1.3 Examples of conjugate prior cases

I used ChatGPT to list out a list of distribution pairs that will result in a conjugate prior case, and I tried to prove each case.

- Binomial likelihood with Beta prior
- Poisson likelihood with Gamma prior
- Gaussian likelihood with Gaussian prior (for the mean μ)
- Gaussian likelihood with Inverse Gamma prior (for the variance σ^2)
- Multinomial likelihood with Dirichlet prior
- Exponential likelihood with Gamma prior

- Geometric likelihood with Beta prior

2 Bionomial likelihood with Beta prior

1, Binomial likelihood w/ Beta Prior $\rightarrow$ WTS Posterior is also Beta

p: prior on success rate : $p \sim Beta(\alpha, \beta)$

X: # successes in n trials : $X|p \sim Binom(n, p)$

$$P(p|X) = \frac{P(X|p)\,P(p)}{\int_p P(X|p)P(p)\,dp} \propto \frac{\binom{n}{x} p^{x}(1-p)^{n-x} \cdot p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

$$= \frac{\binom{n}{x} p^{\alpha-1+x}(1-p)^{\beta-1+n-x}}{B(\alpha,\beta)}$$

$$= \binom{n}{x}\frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)} \frac{p^{\alpha+x-1}(1-p)^{\beta+n-x-1}}{B(\alpha+x,\beta+n-x)}$$

Note about finding normalizing constant for $p(p|X)$

We know that $\int_p P(p|X)\,dp = 1$

$$\Leftrightarrow \underset{\substack{\downarrow \\ \text{Normalizing constant}}}{C} \cdot \binom{n}{x}\frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)} \underbrace{\int_p \underbrace{\frac{p^{\alpha+x-1}(1-p)^{\beta+n-x-1}}{B(\alpha+x,\beta+n-x)}}_{\text{pdf of } p\sim Beta(\alpha+x-1,\beta+n-x)} dp}_{1} = 1$$

$$\Leftrightarrow C = \frac{B(\alpha,\beta)}{B(\alpha+x,\beta+n-x)} \cdot \frac{1}{\binom{n}{x}}$$

So, we proved that

$$P(p|X) = C \cdot \binom{n}{x}\frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)} \frac{p^{\alpha+x-1}(1-p)^{\beta+n-x-1}}{B(\alpha+x,\beta+n-x)}$$

$$= \frac{p^{\alpha+x-1}(1-p)^{\beta+n-x-1}}{B(\alpha+x,\beta+n-x)} \rightarrow \text{pdf of } Beta(\alpha+x,\beta+n-x)$$

$$p|X \sim Beta(\alpha+x, \beta+n-x)$$

3 Poisson likelihood with Gamma prior

2, Poisson Likelihood w/ Gamma Prior

$\lambda \sim \text{Gamma}(\alpha, \beta) \Rightarrow P(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$, where $\underbrace{\Gamma(\overset{\text{integer}}{\alpha}) = (\alpha-1)!}$

↳ Integer def. of gamma funct°

$X \sim \text{Poisson}(\lambda) \Rightarrow P(X|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

$$P(\lambda|x) = \frac{P(x|\lambda)P(\lambda)}{\int_\lambda p(x|\lambda)P(\lambda)\,d\lambda} \propto P(x|\lambda)P(\lambda)$$

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

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

Now, find the normalizing factor for $P(\lambda|x)$ s.t $\int_\lambda P(\lambda|x)\,d\lambda = 1$

$$\Leftrightarrow C \cdot \frac{\beta^\alpha}{x!\,\Gamma(\alpha)} \int_\lambda \lambda^{x+\alpha-1} e^{-\lambda(\beta+1)}\,d\lambda = 1$$

$$\Leftrightarrow C \cdot \frac{\beta^\alpha}{x!\,\Gamma(\alpha)} \frac{\Gamma(x+\alpha)}{(\beta+1)^{\alpha+x}} \underbrace{\int_\lambda \frac{(\beta+1)^{x+\alpha}}{\Gamma(x+\alpha)} \lambda^{x+\alpha-1} e^{-(\beta+1)\lambda}\,d\lambda}_{=1 \text{ b/c that's pdf of } \lambda \sim \text{Gamma}(x+\alpha,\, \beta+1)} = 1$$

$$\Leftrightarrow C = \frac{x!\,\Gamma(\alpha)}{\beta^\alpha} \cdot \frac{(\beta+1)^{\alpha+x}}{\Gamma(x+\alpha)}$$

Therefore $P(\lambda|x) = C \cdot \frac{\beta^\alpha}{x!\,\Gamma(\alpha)} \lambda^{x+\alpha-1} e^{-\lambda(\beta+1)}$

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

$\Rightarrow \lambda|x \sim \text{Gamma}(x+\alpha, \beta+1)$

4 Two cases for the Gaussian distribution, aka more reason why Gaussian runs the world

$\vec{X} \sim N(\vec{\mu}, \Sigma) \qquad \vec{X} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \end{bmatrix}$

Prove that $\qquad \vec{x}_1 | \vec{x}_2 \sim N(\mu_{1|2}, \Sigma_{1|2})$

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

Proof: WLG, we can have $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

1, First, prove the marginal distribution $x_2 \sim N(\mu_2, \Sigma_{22})$

We can prove this by proving that:

$\boxed{\text{Lemma 1}}$ + If $\vec{x} \sim N(\vec{\mu}, \Sigma)$ then any linear transformat° of x is also Normal

$$\underset{m\times 1}{\vec{y}} = \underset{m\times n}{A} \underset{n\times 1}{\vec{x}} + \underset{m\times 1}{\vec{b}} \sim N(\underbrace{A\mu + b}_{m\times 1}, \overbrace{\underset{m\times n}{A} \underset{n\times n}{\Sigma} \underset{n\times m}{A^T}}^{m\times m})$$

(Proof by moment generating function

+ Then if $S = [0 \; 1]$

$$\vec{y} = \underbrace{S\vec{x}}_{x_2} \sim N(\underbrace{S\vec{\mu}}_{[0 \; 1]\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mu_2}, S \Sigma S^T)$$

$$S\Sigma S^T = \underbrace{\underbrace{[0 \; 1] \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}}_{[\sigma_{12} \;\; \sigma_{22}]} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\sigma_{22}}$$

So $x_2 \sim N(\mu_2, \sigma_{22})$

2, Then, given the marginal distribute of a variable derived above, we can prove the conditional distribution

$\boxed{\text{Lemma 2}}$

$$E(x_1 | x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}} (x_2 - \mu_2)$$

$$Var(x_1 | x_2) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$$

3, Now, we can assume:

$$x|\mu \sim \text{Normal}(\mu, \sigma^2)$$
$$\mu \sim \text{Normal}(\mu_0, \sigma_0^2)$$

We have $\begin{bmatrix} x \\ \mu \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} E[x] \\ \mu_0 \end{bmatrix}, \begin{bmatrix} Var(X) & Cov(X,\mu) \\ Cov(X,\mu) & \sigma_0^2 \end{bmatrix}\right)$

+ What's $E[x]$? by previous lemma of conditional dist.

We know $\begin{cases} E[x|\mu] = \mu = E(x) + \dfrac{Cov(X,\mu)}{\sigma_0^2}(\mu - \mu_0) & (\star) \\ Cov(x|\mu) = \sigma^2 = Var(X) - \dfrac{[Cov(X,\mu)]^2}{\sigma_0^2} & (\star\star) \end{cases}$

by previous lemma of conditional dist.

By law of total variance: $Var[X] = E_Y[Var(X|Y)] + Var_Y(E(X|Y))$

So $Var[X] = E_\mu[Var(X|\mu)] + Var_\mu(E(X|\mu))$

$= E_\mu(\sigma^2) + Var_\mu(\mu)$

$Var(X) = \sigma^2 + \sigma_0^2$

Replace into $(\star\star)$:

$$\sigma^2 = \sigma^2 + \sigma_0^2 - \frac{[Cov(X,\mu)]^2}{\sigma_0^2}$$

$$\Leftrightarrow \sigma_0^2 = \frac{[Cov(X,\mu)]^2}{\sigma_0^2}$$

$$\Leftrightarrow Cov(X,\mu) = \sigma_0^2$$

Replace $Cov(X,\mu)$ into $(\star)$

$$\mu = E(x) + \frac{\sigma_0^2}{\sigma_0^2}(\mu - \mu_0)$$

$$\Leftrightarrow E(x) = \mu_0$$

So $\begin{bmatrix} x \\ \mu \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} \mu_0 \\ \mu_0 \end{bmatrix}, \begin{bmatrix} \sigma^2 + \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 \end{bmatrix}\right)$

4, Now, if we apply the rules of conditional distribution from (2)

$\mu|x \sim$ Normal with:

$$E[\mu|x] = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(x - \mu_0)$$

$$Var(\mu|x) = \sigma_0^2 - \frac{\sigma_0^4}{\sigma^2 + \sigma_0^2} = \frac{\sigma_0^2 \sigma^2}{\sigma^2 + \sigma_0^2}$$

5, Can we extend the framework to find $P(\mu|x_1 \dots x_n)$? Yes

But let's turn $P(x_1 \dots x_n|\mu)$ into $P(\bar{x}|\mu)$

$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\Rightarrow \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \sigma^2 I_n \right)$$

By the fact that any linear transformat° of $\vec{X}$ is normal (proved in point (1))

Let $S = [1 \dots 1]_{1\times n}$, then we have:

$$\underbrace{S}_{1\times n} \cdot \underbrace{\vec{X}}_{n\times 1} \sim N\Big(S.\vec{\mu},\ \underbrace{S}_{1\times n} \underbrace{\sigma^2 I_n}_{n\times n} \underbrace{S^T}_{n\times 1}\Big)$$

$$\Leftrightarrow \sum_{i=1}^n X_i \sim N(n.\mu,\ n\sigma^2)$$

$$\Leftrightarrow \frac{\sum_{i=1}^n X_i}{n} \sim N\Big(\mu,\ \frac{\sigma^2}{n}\Big)$$

$\to Var(aX) = a^2 Var(X)$, a is a constant

$$\Leftrightarrow \bar{X} \sim N\Big(\mu,\ \frac{\sigma^2}{n}\Big)$$

Next, We want to show that $P(x_1 \dots x_n|\mu) \propto P(\bar{x}|\mu)$

$$P(x_1 \dots x_n|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Big(\frac{-1}{2\sigma^2}(x_i-\mu)^2\Big)$$

$$\propto_\mu \exp\Big(\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2\Big)$$

$$\propto_\mu \exp\Big(\frac{-1}{2\sigma^2}\Big[\Big(\sum_{i=1}^n x_i^2\Big) - 2\mu\Big(\sum_{i=1}^n x_i\Big) + n\mu^2\Big]\Big)$$

$$\propto_\mu \exp\Big(\frac{-1}{2\sigma^2}(n\mu^2 - 2n\mu.\bar{x})\Big)$$

$$\propto_\mu \exp\Big(\frac{-n}{2\sigma^2}(\bar{x}-\mu)^2\Big)$$

$$\propto_\mu P(\bar{x}|\mu)$$

based on sect°(3) $P(\mu|x_1 \dots x_n) \propto P(x_1 \dots x_n|\mu)\,P(\mu) \propto P(\bar{x}|\mu)P(\mu) \propto P(\mu|\bar{x})$

$\Rightarrow$ $\mu|\bar{x} \sim$ Normal with

$E(\mu|\bar{x}) = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}}(\bar{x} - \mu_0)$ $\quad Var(\mu|\bar{x}) = \frac{\sigma_0^2 \frac{\sigma^2}{n}}{\sigma^2 + \sigma_0^2}$

Prove the following statement

If $x_i \overset{iid}{\sim} N(\mu, \sigma^2)$, where μ is fixed

$\underset{\text{Inverse Gamma}}{\sigma^2 \sim IG(\alpha, \beta)} \Rightarrow P(\sigma^2|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right)$

Then $\sigma^2 | x_1 \ldots x_n \sim IG\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_i (x_i - \mu)^2\right)$

$$P(x_1 \ldots x_n|\sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-1}{2\sigma^2}\sum_i (x_i-\mu)^2\right)$$

$$P(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1}\exp\left(\frac{-\beta}{\sigma^2}\right)$$

$$P(\sigma^2|x_1 \ldots x_n) \propto P(x_1 \ldots x_n|\sigma^2)P(\sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-1}{2\sigma^2}\sum_i (x_i-\mu)^2\right) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1}\exp\left(\frac{-\beta}{\sigma^2}\right)$$

$$\propto_{\sigma} \exp\left(\frac{-1}{\sigma^2}\left(\frac{1}{2}\sum_i (x_i-\mu)^2 + \beta\right)\right)(\sigma^2)^{-\frac{n}{2}-\alpha-1}.$$

Hence $\sigma^2 | x_1 \ldots x_n \sim IG\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_i (x_i-\mu)^2\right)$

Proof of Lemma 1

If $\vec{x} \sim N(\vec{\mu}, \Sigma)$ then any linear transformat° of $\vec{x}$ is also Normal

In particular: $\underbrace{\vec{y}}_{m\times 1} = \underbrace{A}_{m\times n}\underbrace{\vec{x}}_{n\times 1} + \underbrace{\vec{b}}_{m\times 1} \sim N(\underbrace{A\vec{\mu}+\vec{b}}_{m\times 1}, \underbrace{A}_{m\times n}\underbrace{\Sigma}_{n\times n}\underbrace{A^T}_{n\times m})$

Proof: We know that the moment generating function of $\vec{x}$ is:

$$M_x(t) = E(\exp(t^T\vec{x}))$$

Then $M_y(t) = E[\exp(t^T(A\vec{x}+\vec{b}))]$

$$= \exp(t^T\vec{b}) \, . \, E[\exp(t^T A\vec{x})]$$
$$= \exp(t^T\vec{b}) \, . \, M_x(A^T t)$$

The moment generating funct° of $\vec{x} \sim N(\vec{\mu}, \Sigma)$ is

$$M_x(t) = \exp(t^T\mu + \frac{1}{2} t^T\Sigma t)$$

Then, we have: $M_y(t) = \exp(t^T\vec{b}) \, . \, M_x(A^T t)$

$$= \exp(t^T\vec{b}) \, . \, \exp(t^T A\vec{\mu} + \frac{1}{2} t^T A\Sigma A^T t)$$

$$= \exp\left(t^T\underbrace{(\vec{b} + A\vec{\mu})}_{\mu_y} + \frac{1}{2} t^T\underbrace{(A\Sigma A^T)}_{\Sigma_y} t\right)$$

Since MGF and PDF are equivalent, we have

$$y \sim N(\vec{b} + A\vec{\mu}, A\Sigma A^T)$$

Proof of Lemma 2

If $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right)$

Then
$$E(x_1|x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2)$$
$$Var(x_1|x_2) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$$

Proof:
$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} \qquad P(x_2) \longrightarrow x_2 \sim N(\mu_2, \sigma_{22}) \text{ by Lemma 1}$$
$$= \frac{1/\sqrt{(2\pi)^2|\Sigma|} \cdot \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]}{1/\sqrt{2\pi\sigma_{22}} \cdot \exp\left(-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_{22}}\right)}$$

$$[x_1-\mu_1 \quad x_2-\mu_2] \underbrace{\begin{bmatrix} a & b \\ b & c \end{bmatrix}}_{\Sigma^{-1}} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$
$$[a(x_1-\mu_1)+b(x_2-\mu_2) \quad b(x_1-\mu_1)+c(x_2-\mu_2)]$$
$$a(x_1-\mu_1)^2 + 2b(x_1-\mu_1)(x_2-\mu_2) + c(x_2-\mu_2)^2$$

What's a, b, c when $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$

$$\Rightarrow \quad a = \frac{\sigma_{22}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \qquad b = \frac{-\sigma_{12}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \qquad c = \frac{\sigma_{11}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sigma_{22}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}} \cdot \exp\left(-\frac{1}{2}\left(\frac{\sigma_{22}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}(x_1-\mu_1)^2 - \frac{2\sigma_{12}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}(x_1-\mu_1)(x_2-\mu_2) + \frac{\sigma_{11}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}(x_2-\mu_2)^2 - \frac{(x_2-\mu_2)^2}{\sigma_{22}}\right)\right)$$

Define $x_2 - \mu_2 := \hat{\mu}_2$
$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sigma_{22}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}} \cdot \exp\left(-\frac{1}{2} \cdot \frac{\sigma_{22}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\left((x_1-\mu_1)^2 - 2\frac{\sigma_{12}}{\sigma_{22}}(x_1-\mu_1)\hat{\mu}_2 + \frac{\sigma_{11}}{\sigma_{22}}\hat{\mu}_2^2 - \frac{\hat{\mu}_2^2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}{\sigma_{22}^2}\right)\right)$$

Define $\frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}} := \hat{\sigma}^2$

$$= \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \exp\left(-\frac{1}{2}\cdot\frac{1}{\hat{\sigma}^2}\left(x_1^2 - 2\frac{\sigma_{12}}{\sigma_{22}}x_1\hat{\mu}_2 + 2\frac{\sigma_{12}}{\sigma_{22}}\mu_1\hat{\mu}_2 + \mu_1^2 - 2x_1\mu_1 + \frac{\sigma_{11}}{\sigma_{22}}\hat{\mu}_2^2 - \frac{\hat{\mu}_2^2(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}{\sigma_{22}^2}\right)\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \exp\left(\frac{-1}{2\hat{\sigma}^2}\left(x_1^2 - 2x_1\left(\frac{\sigma_{12}}{\sigma_{22}}\hat{\mu}_2 + \mu_1\right) + \frac{\sigma_{11}}{\sigma_{22}}\hat{\mu}_2^2 + 2\frac{\sigma_{12}}{\sigma_{22}}\mu_1\hat{\mu}_2 + \mu_1^2 - \hat{\mu}_2^2\frac{(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}{\sigma_{22}^2}\right)\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \exp\left(\frac{-1}{2\hat{\sigma}^2}\left(x_1^2 - 2x_1\left(\frac{\sigma_{12}}{\sigma_{22}}\hat{\mu}_2 + \mu_1\right) + \mu_1^2 + 2\mu_1\frac{\sigma_{12}}{\sigma_{22}}\hat{\mu}_2 + \hat{\mu}_2^2\left(\underbrace{\frac{\sigma_{11}}{\sigma_{22}} - \frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}^2}}_{\frac{\sigma_{12}^2}{\sigma_{22}^2}}\right)\right)\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \exp\left(\frac{-1}{2\hat{\sigma}^2}\left(x_1^2 - 2x_1\left(\frac{\sigma_{12}}{\sigma_{22}}\hat{\mu}_2 + \mu_1\right) + \left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}\hat{\mu}_2\right)^2\right)\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \exp\left(\frac{-1}{2\hat{\sigma}^2}\left[x_1 - \left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}\hat{\mu}_2\right)\right]^2\right)$$

So $x_1 | x_2 \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(\mu_2 - x_2),\ \frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}}\right)$ ■

5 Multinomial likelihood with Dirichlet prior

4, Multinomial likelihood w/ Dirichlet Prior

$$\vec{p} = (p_1, \dots p_k) \sim \text{Dirichlet}(\vec{\alpha}) \Rightarrow P(\vec{p}) = \frac{1}{B(\vec{\alpha})} \prod_{i=1}^{k} p_i^{\alpha_i - 1}$$

where $B(\vec{\alpha}) = \dfrac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)}$

$$X = (x_1, \dots, x_k) \sim \text{Multinomial}(\vec{p}) \Rightarrow P(\vec{X}) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

$$P(\vec{p} \mid \vec{X}) = \frac{P(\vec{X} \mid \vec{p}) P(\vec{p})}{P(\vec{X})}$$

$$\propto P(\vec{X} \mid \vec{p}) P(\vec{p})$$

$$= C \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \cdot \frac{1}{B(\vec{\alpha})} p_1^{\alpha_1 - 1} \dots p_k^{\alpha_1 - 1}$$

$$= C \cdot \frac{n!}{x_1! \dots x_k!} \cdot \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} p_1^{(x_1 + \alpha_1 - 1)} \dots p_k^{(x_k + \alpha_k - 1)}$$

We want to find the normalizing factor for $P(\vec{p} \mid \vec{X})$ s.t $\int_{\vec{p}} P(\vec{p} \mid \vec{X}) \, d\vec{p}$

$$\Leftrightarrow \quad C \int_{\vec{p}} \frac{n!}{x_1! \dots x_k!} \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} p_1^{(x_1 + \alpha_1 - 1)} \dots p_k^{(x_k + \alpha_k - 1)} \, d\vec{p} = 1$$

$$\Leftrightarrow \quad C \cdot \frac{n!}{x_1! \dots x_k!} \cdot \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \cdot \frac{\prod_{i=1}^{k} \Gamma(x_i + \alpha_i)}{\Gamma(\sum_i x_i + \alpha_i)} \cdot \underbrace{\int_{\vec{p}} \frac{\Gamma(\sum_i x_i + \alpha_i)}{\prod_{i=1}^{k} \Gamma(x_i + \alpha_i)} \cdot p_1^{(x_1 + \alpha_1 - 1)} \dots p_k^{(x_k + \alpha_k - 1)}}_{1} = 1$$

$$\Leftrightarrow \quad C = \frac{x_1! \dots x_k!}{n!} \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)} \cdot \frac{\Gamma(\sum_i x_i + \alpha_i)}{\prod_{i=1}^{k} \Gamma(x_i + \alpha_i)}$$

$$\text{So, } P(\vec{p} \mid \vec{X}) \overset{\text{plugging } c \text{ into } P(\vec{p} \mid \vec{X})}{=} \frac{\Gamma(\sum_i x_i + \alpha_i)}{\prod_{i=1}^{k} \Gamma(x_i + \alpha_i)} \cdot p_1^{(x_1 + \alpha_1 - 1)} \dots p_k^{(x_k + \alpha_k - 1)}$$

$$\vec{p} \mid \vec{X} \sim \text{Dirichlet}(\vec{X} + \vec{p}) \quad \blacksquare$$

6 Exponential likelihood with Gamma prior

5, Exponential likelihood with Gamma Prior

$$\lambda \sim Gamma(\alpha, \beta) \Rightarrow P(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$x \sim Exponential(\lambda) \Rightarrow P(x|\lambda) = \lambda e^{-\lambda x}$$

$$P(\lambda|X) = \frac{P(X|\lambda)P(\lambda)}{P(X)}$$

$$\propto P(X|\lambda)P(\lambda)$$

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

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

We want to find the normalizing factor c s.t. $\int_\lambda P(\lambda|x) = \int_\lambda c. \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^\alpha e^{-\lambda(x+\beta)} = 1$

$$\Leftrightarrow c. \frac{\beta^\alpha}{\Gamma(\alpha)} \left[\int_\lambda \lambda^\alpha e^{-\lambda(\beta+x)} \cdot \frac{(\beta+x)^\alpha}{\Gamma(\alpha+1)} d\lambda \right] \cdot \frac{\Gamma(\alpha+1)}{(\beta+x)^\alpha} = 1$$

$$\Leftrightarrow c = \frac{(\beta+x)^\alpha}{\beta^\alpha} \cdot \frac{\Gamma(\alpha)}{\Gamma(\alpha+1)}$$

replace c

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

Hence $\lambda|x \sim Gamma(\alpha+1, x+\beta)$

7 Geometric likelihood with Beta prior

7) Geometric likelihood w/ Beta Prior

$$p \sim Beta(\alpha, \beta) \Rightarrow P(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \quad \text{where } B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$X_i \overset{iid}{\sim} Geometric(p) \Rightarrow P(\vec{X}) = \prod_i^K (1-p)^{x_i-1} p$$

$$P(p|\vec{X}) \propto P(\vec{X}|p)P(p)$$

$$= (1-p)^{(\Sigma_i x_i)-K} p^K \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \cdot C$$

We want to find C s.t. $\int_p P(p|\vec{X}) = 1$

$$\parallel$$

$$C \cdot \int_p \frac{(1-p)^{(\Sigma_i x_i)-K+\beta-1} p^{K+\alpha-1}}{B(\alpha, \beta)} = 1$$

$$\Leftrightarrow C \cdot \frac{B(K+\alpha, (\Sigma x_i)+\beta-K)}{B(\alpha, \beta)} \underbrace{\int_p \frac{p^{K+\alpha-1}(1-p)^{(\Sigma x_i)-K+\beta-1}}{B(K+\alpha, (\Sigma_i x_i)+\beta-K)}}_{1} = 1$$

$$\Rightarrow C = \frac{B(\alpha, \beta)}{B(\alpha+K, (\Sigma_i x_i)+\beta-K)}$$

So, $$P(p|\vec{X}) = \frac{p^{K+\alpha-1}(1-p)^{(\Sigma_i x_i)-K+\beta-1}}{B(K+\alpha, (\Sigma_i x_i)+\beta-K)}$$

$$\Rightarrow p|\vec{X} \sim Beta(K+\alpha, (\Sigma_i^K x_i)+\beta-K)$$