1. Considera la distribución Exponencial, con función de densidad

$$p(x|\theta) = \theta\, \mathrm{e}^{-\theta x} \quad (0 < x;\ 0 < \theta).$$

(a) Demuestra que la familia Gamma$(\theta|\alpha, \beta)$ es conjugada para θ, exhibiendo la regla de actualización de los hiperparámetros de la distribución de θ dada una muestra aleatoria $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ de $p(x|\theta)$.

(b) Demuestra que $S = \sum_{i=1}^{n} X_i$ es una estadística suficiente para θ utilizando la definición bayesiana de suficiencia.

a) $p(\theta|\hat{x}) \propto p(\theta)\, p(\hat{x}|\theta)$

$$= \left\{ \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \right\} \theta^{n} e^{-\theta \sum_{i=1}^{n} x_i}$$

$$\propto \theta^{\alpha+n-1} e^{-\theta \; \beta + \sum_{i=1}^{n} x_i)}.$$

$\Rightarrow p(\theta|\hat{x}) = \text{Gamma}\left(\alpha+n,\ \beta + \sum_{i=1}^{n} x_i\right)$

b) Como

$$p(\theta|t_n) \propto p(\theta)\, p(t_n|\theta)$$

$$\propto \theta^{\alpha-1} e^{-\beta\theta}\, p(t_n|\theta)$$

$\Rightarrow$ Basta ver que $p(t_n|\theta) \propto \theta^{n} e^{-\theta t_n}$

Pero $p(x_i|\theta) \sim \text{Gamma}(\alpha = 1,\ \beta = \theta)$.

Por lo tanto, por

Theorem 4.6.7 (Generalization of Theorem 4.2.12) *Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $Z = X_1 + \cdots + X_n$. Then the mgf of Z is*

$$M_Z(t) = M_{X_1}(t) \cdot \cdots \cdot M_{X_n}(t).$$

In particular, if $X_1, \ldots, X_n$ all have the same distribution with mgf $M_X(t)$, then

$$M_Z(t) = (M_X(t))^n.$$

Example 4.6.8 (Mgf of a sum of gamma variables) Suppose $X_1, \ldots, X_n$ are mutually independent random variables, and the distribution of X_i is gamma(α_i, β). From Example 2.3.8, the mgf of a gamma(α, β) distribution is $M(t) = (1 - \beta t)^{-\alpha}$. Thus, if $Z = X_1 + \cdots + X_n$, the mgf of Z is

$$M_Z(t) = M_{X_1}(t) \cdot \cdots \cdot M_{X_n}(t) = (1-\beta t)^{-\alpha_1} \cdot \cdots \cdot (1-\beta t)^{-\alpha_n} = (1-\beta t)^{-(\alpha_1 + \cdots + \alpha_n)}.$$

This is the mgf of a gamma$(\alpha_1 + \cdots + \alpha_n, \beta)$ distribution. Thus, the sum of independent gamma random variables that have a common scale parameter β also has a gamma distribution. ||

A generalization of Theorem 4.6.7 is obtained if we consider a sum of linear functions of independent random variables.

obtenemos lo deseado.

2. Considera el modelo Poisson, con función de probabilidad de la forma

$$\mathrm{Po}(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad (x = 0, 1, \ldots; 0 < \lambda).$$

(a) Muestra que este modelo pertenece a una familia exponencial natural, escribiéndolo en su representación canónica.

(b) Encuentra $\mathcal{C}_\theta(\mathcal{F})$, la familia conjugada estándar para el parámetro canónico θ. También encuentra $\mathcal{C}_\lambda^\theta(\mathcal{F})$, la familia conjugada inducida por θ sobre λ.

(c) Demuestra que la familia de distribuciones $\mathrm{Gamma}(\lambda|\alpha, \beta)$ es la familia conjugada estándar para λ, exhibiendo explícitamente la regla de actualización de los hiperparámetros α y β dada una muestra aleatoria $\boldsymbol{X} = (X_1, \ldots, X_n)$ de $\mathrm{Po}(x|\lambda)$. Denota a esta familia por $\mathcal{C}_\lambda(\mathcal{F})$, como lo hicimos en clase.

(d) Demuestra que $\mathcal{C}_\lambda(\mathcal{F}) = \mathcal{C}_\lambda^\theta(\mathcal{F})$.

a) $$\frac{1}{x!} \lambda^x \exp(-\lambda) = \frac{1}{x!} \exp(x \log \lambda) \exp(-\lambda)$$

$$= \frac{1}{x!} \exp(x \log \lambda - \lambda)$$

$$\Longrightarrow \quad a(x) = \frac{1}{x!} \mathbb{1}_{\mathbb{N}}(x), \quad \theta = \log \lambda, \quad t(x) = x$$

Por lo tanto

$$M(\theta) = \log \int \frac{1}{x!} \mathbb{1}_{\mathbb{N}}(x) \exp(x \log \lambda)\, dx$$

$$= \log \sum_{x=0}^{\infty} \frac{1}{x!} \lambda^x = \log e^\lambda = \lambda \quad \checkmark$$

b) Los elementos de $C_\theta(\mathcal{F})$ son de la forma

$$p_\theta(\theta | s_0, n_0) = h(s_0, n_0) \exp\{ s_0 \theta - n_0 \overbrace{M(\theta)}^{\exp(\theta)} \}. \quad (\text{cf p. 52})$$

$\Longrightarrow$ Como $\theta = \log \lambda$ &

$$p_\lambda^{\theta}(\lambda \mid s_0, n_0) = p_\theta(\theta(\lambda) \mid s_0, n_0) \, | J_\theta(\lambda) |$$

$$= p_\theta(\log \lambda \mid s_0, n_0) \frac{1}{\lambda}$$

$$\propto \exp\{ s_0 \log \lambda - n_0 \lambda \} \frac{1}{\lambda}$$

$$= \exp\{ \log(\lambda^{s_0}) \} \exp\{ -n_0 \lambda \} \frac{1}{\lambda}$$

$$= \lambda^{s_0} \exp\{ -n_0 \lambda \} \lambda$$

$$= \lambda^{s_0 - 1} \exp\{ -n_0 \lambda \}$$

Así son los elementos de $C_\lambda^{\theta}(\mathcal{F})$

c) También tenemos que $C_\lambda(\mathcal{F})$ tiene

$p_\lambda(\lambda \mid s_0, n_0) \propto \lambda^{s_0} e^{-n_0 \lambda}$ (cf p60)

Poniendo $s_0 = \alpha - 1$ & $n_0 = \beta$ es claro que

$C_\lambda(\mathcal{F})$ es la familia $\mathrm{Gam}(\lambda \mid \alpha, \beta)$

d) Es consecuencia inmediata de b y c.

3. Sea $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ una muestra aleatoria de observaciones i.i.d. de una distribución Poisson, con función de probabilidad

$$\Pr[X = x|\lambda] = \frac{\lambda^x e^{-\lambda}}{x!} \qquad (x = 0, 1, \ldots; 0 < \lambda).$$

Sea X_* el valor de una observación futura de este modelo.

(a) Suponiendo una distribución inicial Gamma$(\lambda|\alpha, \beta)$ para λ, encuentra el valor esperado de la distribución final del parámetro $\phi = \phi(\lambda) = \Pr[X_* = 0|\lambda]$.

(b) Encuentra la distribución predictiva final de X_*, dada por

$$\Pr[X_* = x_*|\boldsymbol{x}] = \int \Pr[X_* = x_*|\lambda]\, p(\lambda|\boldsymbol{x})\, d\lambda,$$

donde $p(\lambda|\boldsymbol{x})$ denota a la función de densidad de la distribución final de λ. En particular, ¿cuál es el valor de $\Pr[X_* = 0|\boldsymbol{x}]$?

(c) Compara (a) con (b). ¿Qué puedes decir sobre la probabilidad de que X_* sea igual a cero en cada caso?

a) $\mathbb{E}[\phi \mid x] = \int \phi \, p(\phi \mid x) \, d\phi$

$= \int \Pr[X_* = 0 \mid x] \, p(\phi \mid x) \, d\phi$

$= \int e^{-\lambda} \, p(\phi \mid x) \, d\phi$

$= e^{-\lambda}$

b) Primero calculamos

$$p(\lambda|\hat{x}) = \frac{p(\lambda)\, p(\hat{x}|\lambda)}{\int p(\lambda)\, p(\hat{x}|\lambda)\, d\lambda}$$

$$= \frac{\mathrm{Gamma}(\lambda|\alpha,\beta)\, \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{x_1! \cdots x_n!}}{\int \mathrm{Gamma}(\lambda|\alpha,\beta)\, \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{x_1! \cdots x_n!}\, d\lambda}$$

$$= \frac{\cancel{\frac{\beta^\alpha}{\Gamma(\alpha)}}\, \lambda^{\alpha-1} e^{-\beta\lambda}\, \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\cancel{x_1! \cdots x_n!}}}{\int \cancel{\frac{\beta^\alpha}{\Gamma(\alpha)}}\, \lambda^{\alpha-1} e^{-\beta\lambda}\, \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\cancel{x_1! \cdots x_n!}}\, d\lambda}$$

$$= \frac{\lambda^{\alpha-1+\sum_{i=1}^n x_i} e^{-(\beta+n)\lambda}}{\int \lambda^{\alpha-1+\sum_{i=1}^n x_i} e^{-(\beta+n)\lambda}\, d\lambda}$$

$$= \frac{(\beta+n)^{\alpha+\sum_{i=1}^n x_i}}{\Gamma(\alpha+\sum_{i=1}^n x_i)}\, \lambda^{\alpha-1+\sum_{i=1}^n x_i} e^{-(\beta+n)\lambda}$$

$$= \mathrm{Gamma}(\lambda|\alpha+\textstyle\sum x_i, \beta+n)$$

Ahora sí,

$$\int \Pr[X_* = x_* \mid \lambda]\, p(\lambda \mid \hat{x})\, d\lambda =$$

$$\int \frac{\lambda^{x_*} e^{-\lambda}}{x_*!} \frac{(\beta+n)^{\alpha+\sum_{i=1}^n x_i}}{\Gamma(\alpha+\sum_{i=1}^n x_i)} \lambda^{\alpha-1+\sum_{i=1}^n x_i} e^{-(\beta+n)\lambda}\, d\lambda =$$

$$\frac{1}{x_*!} \frac{(\beta+n)^{\alpha+\sum_{i=1}^n x_i}}{\Gamma(\alpha+\sum_{i=1}^n x_i)} \int \lambda^{\alpha+x_*+\sum_{i=1}^n x_i - 1} e^{-(\beta+n+1)\lambda}\, d\lambda =$$

$$\frac{1}{x_*!} \frac{(\beta+n)^{\alpha+\sum_{i=1}^n x_i}}{\Gamma(\alpha+\sum_{i=1}^n x_i)} \frac{\Gamma(\alpha+x_*+\sum_{i=1}^n x_i)}{(\beta+n+1)^{\alpha+x_*+\sum_{i=1}^n x_i}} =$$

$$\frac{(\beta+n)^{\alpha+\sum_{i=1}^n x_i}}{(\beta+n+1)^{\alpha+x_*+\sum_{i=1}^n x_i}}.$$

Ocupamos que $\frac{\Gamma(z+1)}{\Gamma(z)} = z$

4. Desarrolla el ejemplo relativo a la distribución normal que vimos en clase (páginas 66 y 67 de las notas "Inferencia.pdf"). Específicamente,

(a) Muestra que la familia Normal/Gamma-inversa es conjugada para el modelo $N(x|\mu, \sigma^2)$, verificando la regla de actualización de los hiperparámetros vista en clase.

(b) Muestra que la distribución final marginal de μ es t de Student,

$$\mathrm{St}(\mu | m_x, n_x a_x / b_x, a_x).$$

(c) Adicionalmente encuentra $p(y|\boldsymbol{x})$, la distribución predictiva final de la observación futura $Y = X_{n+1}$ para este modelo.

a) Primero veamos $p(\mu | \sigma^2, x) = N(\mu \,|\, m_x, \sigma^2/n_x)$.

Por Bayes,

$$p(\mu | \sigma^2, x) \propto p(x | \mu, \sigma^2)\, p(\mu | \sigma^2)$$

$$= \left\{ \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x_i - \mu)^2}{2\sigma^2} \right) \right\} \times \frac{1}{\sqrt{2\pi(\sigma^2/n_0)}} \exp\left(- \frac{(\mu - m_0)^2}{2\frac{\sigma^2}{n_0}} \right)$$

$$\propto \exp\left\{ - \frac{(\mu - m_0)^2}{2\sigma^2/n_0} - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right\}$$

$$\propto \exp\left\{ - \frac{1}{2\sigma^2} \left[n_0 (\mu - m_0)^2 + \sum_{i=1}^{n} (x_i - \mu)^2 \right] \right\}$$

$$\propto \exp\left\{ - \frac{1}{2\sigma^2} \left[n_0 \mu^2 - 2 n_0 \mu m_0 + \sum_{i=1}^{n} (\mu^2 - 2\mu x_i) \right] \right\}$$

Los sumandos que no dependen de μ salen como factores constantes de la forma exp(algo que no depende de μ).

$$\propto \exp\left\{ - \frac{1}{2\sigma^2} \left[n_0 \mu^2 - 2 n_0 \mu m_0 + n\mu^2 + 2\mu n \bar{x} \right] \right\} \quad \cdots (1)$$

Por otro lado,

$$N(\mu \mid m_x, \sigma^2/n_x)$$

$$\propto \exp\left\{-\frac{n_x(\mu-m_x)^2}{2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{1}{2\sigma^2}(n_0+n)\left(\mu-\frac{n_0m_0+n\bar{x}}{n_0+n}\right)^2\right\}$$

$$\propto \exp\left\{\frac{1}{2\sigma^2}\frac{\left(\mu(n_0+n)-n_0m_0+n\bar{x}\right)^2}{n_0+n}\right\}$$

$$\propto \exp\left\{\frac{1}{2\sigma^2}\left[\frac{\mu^2(n_0+n)^{\cancel{2}}}{\cancel{n_0+n}}-\frac{2\mu\cancel{(n_0+n)}n_0m_0}{\cancel{n_0+n}}+\frac{2\mu\cancel{(n_0+n)}n\bar{x}}{\cancel{n_0+n}}\right]\right\}$$

$$\propto \exp\left\{\frac{1}{2\sigma^2}\left[\mu^2(n_0+n)-2\mu n_0m_0+2\mu n\bar{x}\right.\right\}$$

Comparando esto con (1) vemos que coinciden y por lo tanto,

$$p(\mu \mid \sigma^2, x) = N(\mu \mid m_x, \sigma^2/n_x).$$

Ahora veamos que $p(\sigma^2 \mid x) = IGa(\sigma^2 \mid a_x/2, b_x/2)$.

Primero notemos que

$$p(x \mid \sigma^2) = \frac{p(x, \sigma^2)}{p(\sigma^2)} = \frac{\int p(x, \sigma^2, \mu)\, d\mu}{p(\sigma^2)}$$

$$= \frac{1}{p(\sigma^2)} \int p(x \mid \sigma^2, \mu)\, p(\sigma^2, \mu)\, d\mu$$

Por lo tanto,

$$p(\sigma^2 \mid x) \propto p(\sigma^2)\, p(x \mid \sigma^2)$$

$$\propto \int p(x \mid \sigma^2, \mu)\, p(\sigma^2, \mu)\, d\mu$$

$$= \int N(x \mid \mu, \sigma^2)\, N(\mu \mid m_0, \sigma^2/n_0)\, IGa(\sigma^2 \mid a_0/2, b_0/2)\, d\mu$$

$$= \int \Bigg[\prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ - \frac{(x_i - \mu)^2}{2\sigma^2} \right\} \right)$$

$$\times \frac{1}{\sqrt{2\pi\, \sigma^2/n_0}} \exp\left\{ - \frac{(\mu - m_0)^2}{2\, \sigma^2/n_0} \right\}$$

$$\times \frac{(b_0/2)^{a_0/2}}{\Gamma(a_0/2)} (1/\sigma^2)^{a_0/2 + 1} \exp(b_0/2) \Bigg] d\mu$$