6.434/16.391 Statistics for Engineers and Scientists

Lecture 2 09/10/2012

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 2 09/10/2012

DISTRIBUTION THEORY FOR TRANSFORMATION

Topic

- Distribution theory for transformation of random variable (r.v's)
  - Check chapter 2 of Casella & Berger

- Distribution theory for transformation of Random Vectors (R.V's)
  - There are many instant in engineering statistic that we need to know functions of a random sample
  - e.g., sum $X_1 + X_2 + \cdots + X_N$

    sum of square $X_1^2 + X_2^2 + \cdots + X_N^2$

    averages $\frac{1}{N}\sum_{i=1}^{N} X_i$ and $\frac{1}{N}\sum_{i=1}^{N} X_i^2$

Transformation & Jacobian

- Let $\mathbf{h} : \mathbb{R}^K \to \mathbb{R}^K$, $\mathbf{h} = \left[h_1 \; h_2 \; \cdots \; h_K\right]^{\mathrm{T}}$, $h_i : \mathbb{R}^K \to \mathbb{R}$
  $\mathbf{h}$ is called a transformation from $\mathbb{R}^K$ to $\mathbb{R}^K$

- The Jacobian $J_{\mathbf{h}}(\mathbf{t})$ of $\mathbf{h}$ evaluated at $\mathbf{t} = \left[t_1 \; t_2 \; \cdots \; t_K\right]^{\mathrm{T}}$ is

$$J_{\mathbf{h}}(\mathbf{t}) = \det\left(\begin{bmatrix} \frac{\partial h_1(\mathbf{t})}{\partial t_1} & \frac{\partial h_2(\mathbf{t})}{\partial t_1} & \cdots & \frac{\partial h_K(\mathbf{t})}{\partial t_1} \\ \frac{\partial h_1(\mathbf{t})}{\partial t_2} & \frac{\partial h_2(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial h_K(\mathbf{t})}{\partial t_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_1(\mathbf{t})}{\partial t_K} & \frac{\partial h_2(\mathbf{t})}{\partial t_K} & \cdots & \frac{\partial h_K(\mathbf{t})}{\partial t_K} \end{bmatrix}\right)$$

Distribution theory for transformation

- Theorem: Let $\mathbf{X}$ be continuous random variable, $\mathbb{S}$ be open subset of $\mathbb{R}^K$ such that $P(\mathbf{X} \in \mathbb{S}) = 1$. If $\mathbf{g} = [g_1 \; g_2 \; \cdots \; g_K]^{\mathrm{T}}$ is a transformation from $\mathbb{S}$ to $\mathbb{R}^K$, such that
  - $\mathbf{g}$ has continuous first partial derivate in $\mathbb{S}$
  - $\mathbf{g}$ is one to one mapping on $\mathbb{S}$ ($\mathbf{g}(\mathbf{x}) \neq \mathbf{g}(\mathbf{y})$ for $\mathbf{x} \neq \mathbf{y}$)
  - The Jacobian of $\mathbf{g}$ does not vanish on $\mathbb{S}$

  Then the pdf of $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ is given by

$$P_{\mathbf{Y}}(\mathbf{y}) = P_{\mathbf{X}}\left(\mathbf{g}^{-1}(\mathbf{y})\right)\left|J_{\mathbf{g}^{-1}}(\mathbf{y})\right|$$

  where

$$J_{\mathbf{g}^{-1}}(\mathbf{y}) = \frac{1}{J_{\mathbf{g}}\left(\mathbf{g}^{-1}(\mathbf{y})\right)}$$

Example 1 (1 of 2)

- Consider $\mathbf{X} = [X_1 \ X_2]^{\mathrm{T}}$, where $X_1 \sim N(0,\,1)$ and $X_2 \sim N(0,\,4)$ are independent, i.e.,

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \frac{1}{\sqrt{2\pi 4}} e^{-\frac{1}{8}x_1^2} = \frac{1}{4\pi} e^{-\frac{1}{2}\left(x_1^2 + \frac{1}{4}x_2^2\right)}$$

  Let $\mathbf{Y} = [Y_1 \ Y_2]^{\mathrm{T}}$ and $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, where $\mathbf{g} = [g_1 \ g_2]^{\mathrm{T}}$ is given by

$$g_1(\mathbf{X}) = X_1 + X_2 \ \text{ and } \ g_2(\mathbf{X}) = X_1 - X_2$$

  Then we have $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$, where $\mathbf{g}^{-1} = \left[g_1^{-1} \ g_2^{-1}\right]^{\mathrm{T}}$ is given by

$$g_1^{-1}(\mathbf{Y}) = \frac{Y_1 + Y_2}{2} \ \text{ and } \ g_2^{-1}(\mathbf{Y}) = \frac{Y_1 - Y_2}{2}$$

  and we have

$$J_{\mathbf{g}}(\mathbf{x}) = \det\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right) = -2 \qquad J_{\mathbf{g}^{-1}}(\mathbf{y}) = \det\left(\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}\right) = -1/2$$

Example 1 (2 of 2)

- Finally,

$$
\begin{aligned}
f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1,X_2}\left(\frac{Y_1+Y_2}{2},\frac{Y_1-Y_2}{2}\right)\times\frac{1}{2} \\
&= \frac{1}{4\pi}e^{-\frac{1}{2}\left[\frac{1}{4}(y_1+y_2)^2+\frac{1}{16}(y_1-y_2)^2\right]}\times\frac{1}{2} \\
&= \frac{1}{8\pi}e^{-\frac{1}{32}\left(5y_1^2+5y_2^2+6y_1y_2\right)}
\end{aligned}
$$

Example 2 (1 of 2)

- Let $\mathbf{X} = [X_1\ X_2]^{\mathrm{T}}$ for $X_1, X_2 \sim N(0,\,1)$ and X_1 is independent of X_2, i.e.,

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}x_1^2} e^{-\frac{1}{2}x_2^2}$$

Let $\mathbf{Y} = [Y_1\ Y_2]^{\mathrm{T}}$ and $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, where $\mathbf{g} = [g_1\ g_2]^{\mathrm{T}}$ is given by

$$g_1(\mathbf{X}) = X_1 + X_2 \ \text{ and } \ g_2(\mathbf{X}) = X_2 + 1$$

Then we have $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$, where $\mathbf{g}^{-1} = \left[g_1^{-1}\ g_2^{-1}\right]^{\mathrm{T}}$ is given by

$$g_1^{-1}(\mathbf{Y}) = Y_1 - Y_2 + 1 \ \text{ and } \ g_2^{-1}(\mathbf{Y}) = Y_2 - 1$$

and we have

$$J_{\mathbf{g}}(\mathbf{x}) = \det\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\right) \qquad J_{\mathbf{g}^{-1}}(\mathbf{y}) = \det\left(\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\right)$$

$$= 1 \qquad\qquad = 1$$

Example 2 (2 of 2)

- Finally,

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1,X_2}(y_1 - y_2 + 1, y_2 - 1) \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(y_1 - y_2 + 1)^2} e^{-\frac{1}{2}(y_2 - 1)^2} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}\left(y_1^2 + 2y_2^2 + 2y_1 y_2 + 2y_1 - 4y_2 + 2\right)} \end{aligned}$$

Special case: affine transformation

- $\mathbf{g}$ is affine transformation of $\mathbb{R}^K$, i.e.,

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}) = \mathbf{A}\mathbf{X} + \mathbf{c}$$

  where $\mathbf{A} \in \mathbb{R}^{K \times K}$ and $\mathbf{c} \in \mathbb{R}^K$
  - If $\mathbf{c} = \mathbf{0}$, $\mathbf{g}$ is linear, i.e., $\mathbf{g}(\mathbf{X}) = \mathbf{A}\mathbf{X}$

- If $\mathbf{A}$ is non-singular, $\mathbf{g}$ is one-to-one mapping, and

$$\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y}) = \mathbf{A}^{-1}(\mathbf{Y} - \mathbf{c})$$

  In this case, $J_{\mathbf{g}^{-1}}(\mathbf{y}) = \left|\mathbf{A}^{-1}\right| = \frac{1}{|\mathbf{A}|}$, and $J_{\mathbf{g}}(\mathbf{X}) = |\mathbf{A}|$

Lecture 2 09/10/2012

SPECIAL DISTRIBUTIONS

Beta function

- For $\alpha > 0$ and $\beta > 0$ (not necessarily integer), the beta function is given by

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$= 2\int_0^{\frac{\pi}{2}} (\sin\theta)^{2\alpha-1} (\cos\theta)^{2\beta-1}\, d\theta$$

where $x = \sin^2\theta$

Gamma function

- For $\alpha > 0$ (not necessarily integer), the gamma function is given by

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty t^{\alpha-1} e^{-t} dt \\ &= 2\int_0^\infty y^{2\alpha-1} e^{-y^2} dy\end{aligned}$$

where $t = y^2$. Gamma function has following properties

  - $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
  - $\Gamma(0.5) = \sqrt{\pi}$
  - $\Gamma(n) = (n-1)!$ for $n \in \mathbb{Z}^+$
  - Relationship to beta function

$$B(\alpha,\, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Beta and Gamma distribution

- Definition: Random variable X has beta distribution if its pdf is given by

$$f(x|\alpha,\,\beta) = \begin{cases} \frac{1}{B(\alpha,\,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

- Definition: Random variable X has gamma distribution if its pdf is given by

$$f(x|\alpha,\,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

  - If $\alpha = \frac{n}{2}$ and $\beta = 2$, $\text{gamma}\left(\frac{n}{2},\, 2\right)$ is chi-squared distribution with n degrees of freedom
  - If $\alpha = 1$, $\text{gamma}\,(1,\,\beta)$ becomes exponential distribution with mean β

Relationship between Gaussian and Gamma

- Theorem: If $Z \sim N(0,\, 1)$, then $X = Z^2 \sim \mathrm{gamma}\left(\frac{1}{2},\, 2\right)$

- Proof: First calculate the CDF of X

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(Z^2 \leq x\right) \\ &= P\left(-\sqrt{x} \leq Z \leq \sqrt{x}\right) \\ &= F_Z\left(\sqrt{x}\right) - F_Z\left(-\sqrt{x}\right) \\ &= F_Z\left(\sqrt{x}\right) - \left(1 - F_Z\left(\sqrt{x}\right)\right) \\ &= 2F_Z\left(\sqrt{x}\right) - 1 \end{aligned}$$

then

$$f_X(x) = \frac{\partial}{\partial x} F_X(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$$

Functions of Gamma (1 of 3)

- Theorem: Let $X_1 \sim \mathrm{gamma}(\alpha_1,\, \beta)$ and $X_2 \sim \mathrm{gamma}(\alpha_2,\, \beta)$ are two independent random variables with gamma distribution. Then
  - a) $Y_1 = X_1 + X_2 \sim \mathrm{gamma}(\alpha_1 + \alpha_2,\, \beta)$
  - b) $Y_2 = \frac{X_1}{X_1 + X_2} \sim \mathrm{beta}(\alpha_1,\, \alpha_2)$
  - c) Y_1 and Y_2 are independent

Functions of Gamma (2 of 3)

- Proof: Using distribution theory for transformations. Since X_1 and X_2 are independent, their joint distribution is given by

$$f_{X_1,X_2}(x_1,\, x_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} x_1^{\alpha_1-1} x_2^{\alpha_2-1} e^{-(x_1+x_2)/\beta}$$

Let $\mathbf{Y} = [Y_1\ Y_2]^{\mathrm{T}}$ and $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, where $\mathbf{g} = [g_1\ g_2]^{\mathrm{T}}$ is given by

$$g_1(\mathbf{X}) = X_1 + X_2 \ \text{ and } \ g_2(\mathbf{X}) = \tfrac{X_1}{X_1+X_2}$$

Then we have $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$, where $\mathbf{g}^{-1} = \left[g_1^{-1}\ g_2^{-1}\right]^{\mathrm{T}}$ is given by

$$g_1^{-1}(\mathbf{Y}) = Y_1 Y_2 \ \text{ and } \ g_2^{-1}(\mathbf{Y}) = Y_1(1-Y_2)$$

The Jacobian

$$J_{\mathbf{g}^{-1}}(\mathbf{y}) = \det\left(\begin{bmatrix} y_2 & 1-y_2 \\ y_1 & -y_1 \end{bmatrix}\right) = -y_1$$

Functions of Gamma (3 of 3)

- Finally, we have

$$
\begin{aligned}
f_{Y_1,Y_2}(y_1,\, y_2) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}}(y_1 y_2)^{\alpha_1-1}\left[y_1(1-y_2)\right]^{\alpha_2-1} e^{-y_1/\beta} y_1 \\
&= \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1+\alpha_2-1} e^{-y_1/\beta} \\
&\quad \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_2^{\alpha_1-1}(1-y_2)^{\alpha_2-1}
\end{aligned}
$$

- Therefore
  - a) $Y_1 = X_1 + X_2 \sim \mathrm{gamma}(\alpha_1+\alpha_2,\, \beta)$
  - b) $Y_2 = \frac{X_1}{X_1+X_2} \sim \mathrm{beta}(\alpha_1,\, \alpha_2)$
  - c) Y_1 and Y_2 are independent

Exponential Family

- Definition: A family of pdfs or pmfs is called exponential family if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right)$$

where $h(\mathbf{x}) \geq 0$; $t_1(x), t_2(x), \ldots, t_k(x)$ are real valued function of observation x; $c(\boldsymbol{\theta}) \geq 0$; and $w_1(\boldsymbol{\theta}), w_2(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are real valued function of possibly vector-valued $\boldsymbol{\theta}$