6.434/16.391 Statistics for Engineers and Scientists

Lecture 3 09/12/2012

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 3 09/12/2012

DISTRIBUTION OF SAMPLE MEAN AND VARIANCE

Population mean and variance

- Population of size N is given by the following set

$$\{x_1,\ x_2,\ \ldots,\ x_N\}$$

- Population mean: $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

- Population variance: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$

Sample mean and variance

- The sample of size n , $\{x_1, x_2, \ldots, x_n\}$, is a subset of the population
- Sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
  - Intuition for $n-1$: Since $\sum_{i=1}^{n}(x_i - \bar{x})^2$ tend to be smaller than $\sum_{i=1}^{n}(x_i - \mu)^2$, we divide by $n-1$ to compensate
- We will derive the distributions for sample mean and sample variance, based on the following facts

Fact 1 (1 of 2)

- If $Z \sim N(0,\,1)$, then $Z^2 \sim \mathcal{X}_1^2$
  - Proof idea: $P\left(Z^2 \leq z\right) = P\left(-\sqrt{z} \leq Z \leq \sqrt{z}\right)$

- Corollary: If $Z_1,\, Z_2,\, \ldots\, Z_n$ are i.i.d. with $Z_i \sim N(0,\,1)$, then

$$Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \mathcal{X}_n^2$$

  - Proof idea 1: use moment generating function
  - Proof idea 2: use the fact in the next slide [Bickel & Doksum, Vo. I, P489]

Fact 1 (2 of 2)

- [Bickel & Doksum, Vo. I, P489] If $X_1 \sim \text{gamma}(\alpha_1, \beta)$ and $X_2 \sim \text{gamma}(\alpha_2, \beta)$ are independent, then
  - $Y_1 = X_1 + X_2 \sim \text{gamma}(\alpha_1 + \alpha_2, \beta)$
  - $Y_2 = \frac{X_1}{X_1 + X_2} \sim \text{beta}(\alpha_1, \alpha_2)$
  - Y_1 and Y_2 are independent

- Note: Chi-squared distribution with n degrees of freedom is a special type of gamma with $\alpha = n/2$ and $\beta = 2$

Fact 2 (1 of 2)

- Suppose

$$\mathbf{X} = [X_1\ X_2\ \cdots\ X_n]^{\mathrm{T}} \qquad X_i \sim N(\mu_x,\, \sigma_x^2)$$

$$\mathbf{Y} = [Y_1\ Y_2\ \cdots\ Y_m]^{\mathrm{T}}, \qquad Y_i \sim N(\mu_y,\, \sigma_y^2)$$

one way to compare $\mathbf{X}$ and $\mathbf{Y}$ is to look at

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\frac{1}{m}\sum_{i=1}^{m} Y_i^2}$$

Fact 2 (2 of 2)

- Corollary: Let $V \sim \mathcal{X}_n^2$ and $W \sim \mathcal{X}_m^2$ be independent gamma distributed random variables. Then

$$\frac{V/n}{W/m}$$

has F distribution with n and m degrees of freedom

Fact 3

- Corollary: Let $X_1, X_2, \ldots, X_n$ be samples from $N(0, \sigma^2)$, i.e., $X_i \sim N(0, \sigma^2)$. Then

$$\frac{\sum_{i=1}^{k} X_i^2/k}{\sum_{i=k+1}^{k+m} X_i^2/m} \sim F_{k,m}$$

  Let $Z \sim N(0, 1)$ and $V \sim \mathcal{X}_n^2$. Z and V are independent. Then

$$Q = \frac{Z}{\sqrt{V/n}}$$

  has t distribution with n degrees of freedom, i.e.,

$$Q \sim t_n$$

Fact 4: Orthogonal transformation of joint Gaussian

- Theorem: If $\mathbf{Z} = [Z_1\ Z_2\ \cdots\ Z_n]^{\mathrm{T}}$ has independent normally distributed elements with $\mathrm{E}\mathbf{Z} = \mathbf{d}$ and the same variance σ^2 . Let

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

be affine transformation with orthogonal matrix $\mathbf{A}$ and vector $\mathbf{c} = [c_1\ c_2\ \cdots\ c_n]^{\mathrm{T}}$. Then $\mathbf{Y}$ has independent normally distributed components with the same variance σ^2, and

$$\mathrm{E}\mathbf{Y} = \mathbf{A}\mathbf{d} + \mathbf{c}$$

Fact 4: Proof (1 of 2)

- Using distribution theory for transformation,

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Z}}\left(\mathbf{A}^{-1}(\mathbf{Y}-\mathbf{c})\right)\frac{1}{|\det \mathbf{A}|}$$

Since $\mathbf{A}$ is orthogonal, i.e., $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I}$, we have $\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}}$ and $|\det \mathbf{A}| = 1$. Thus,

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Z}}\left(\mathbf{A}^{\mathrm{T}}(\mathbf{Y}-\mathbf{c})\right)$$

- The pdf of $\mathbf{Z}$ is given by

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i-d_i)^2\right\} \\ &= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n}\exp\left\{-\frac{1}{2\sigma^2}|\mathbf{z}-\mathbf{d}|^2\right\} \end{aligned}$$

where $|\mathbf{z}-\mathbf{d}|^2 = \sum_{i=1}^{n}(z_i-d_i)^2$

Fact 4: Proof (2 of 2)

- Substituting $\mathbf{z} = \mathbf{A}^{\mathrm{T}}(\mathbf{y} - \mathbf{c})$, we have

$$\begin{aligned} \left|\mathbf{z} - \mathbf{d}\right|^2 &= \left|\mathbf{A}^{\mathrm{T}}\mathbf{y} - \mathbf{A}^{\mathrm{T}}\mathbf{c} - \mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{d}\right|^2 \\ &= \left|\mathbf{A}^{\mathrm{T}}\left[\mathbf{y} - (\mathbf{A}\mathbf{d} + \mathbf{c})\right]\right|^2 \\ &= \left|\mathbf{y} - (\mathbf{A}\mathbf{d} + \mathbf{c})\right|^2 \end{aligned}$$

  Finally

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left\{-\frac{1}{2\sigma^2}\left|\mathbf{y} - (\mathbf{A}\mathbf{d} + \mathbf{c})\right|^2\right\}$$

- Note: Z_i 's do not have to be identically distributed. They can have different means, i.e., $\mathrm{E}Z_i = d_i$

Distribution of sample mean and variance

- Theorem: Let $\mathbf{Z} = [Z_1 \; Z_2 \; \cdots \; Z_n]^{\mathrm{T}}$ be a random sample from a $N(\mu, \sigma^2)$ population. Then

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

and

$$W = \sum_{i=1}^{n} (Z_i - \bar{Z})^2$$

are independent. Furthermore,

$$\bar{Z} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{W}{\sigma^2} \sim \mathcal{X}_{n-1}^2$$

Proof (1 of 4)

- First of all, find an orthogonal matrix $\mathbf{A}$ such that for $\mathbf{Y} = \mathbf{AZ}$ we have $Y_1 = \sqrt{n}\bar{Z}$ and $\sum_{i=2}^{n} Y_i^2 = W$

- Orthogonal matrix $\mathbf{A}$ is constructed as

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

where the rows are obtained through the "Gram-Schmidt" orthonormalization procedure

Proof (2 of 4)

- Let $\mathbf{Y} = \mathbf{AZ}$. We have

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i = \sqrt{n}\bar{Z}$$

Meanwhile,

$$\sum_{i=1}^{n} Y_i^2 = |\mathbf{Y}|^2 = |\mathbf{AZ}|^2 = |\mathbf{Z}|^2 = \sum_{i=1}^{n} Z_i^2$$

Then

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} Z_i^2 - Y_1^2 = \sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2 = \sum_{i=1}^{n} \left(Z_i - \bar{Z}\right)^2$$

where the last equality is due to

$$\sum_{i=1}^{n} \left(Z_i - \bar{Z}\right)^2 = \sum_{i=1}^{n} \left(Z_i^2 - 2Z_i\bar{Z} + \bar{Z}^2\right)$$

$$= \sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2$$

Proof (3 of 4)

- Using the previous theorem on orthogonal transformation of joint Gaussian, we have

$$\bar{Z} = \frac{Y_1}{\sqrt{n}} \sim N(\mu,\ \frac{\sigma^2}{n})$$

and $\bar{Z}$ is independent of W, which is because Y_1 is independent of $\sum_{i=2}^{n} Y_i^2$

- Further, Y_i 's are independent normally distributed random variable with variance σ^2, and for $i > 1$,

$$\mathrm{E}Y_i = \sum_{j=1}^{n} a_{ij} \mathrm{E} Z_j = \mu \sum_{j=1}^{n} a_{ij} = 0$$

where the last equality is due to the fact that $\mathbf{A}$ is orthogonal matrix, which implies

$$\sum_{j}^{n} \quad \overline{\ \ n} \quad ij \qquad \overline{\ \ n} \sum_{j}^{n} \qquad ij$$

Proof (4 of 4)

- Therefore, $Y_i \sim N(0,\, \sigma^2)$ for $i > 1$, and $\frac{Y_i}{\sigma} \sim N(0,\, 1)$. Thus

$$\frac{W}{\sigma^2} = \sum_{i=2}^{n} \left(\frac{Y_i}{\sigma} \right)^2 \sim \mathcal{X}^2_{n-1}$$