Lab 8

Sampling Theory & Multivariate Transformations

8.1. Normal Distribution Theory

Theorem 8.1.1. Suppose $\mathbf{X} = (X_1, \cdots, X_n) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The components of $\mathbf{X}$ are independent if and only if $\rho_{ij} = 0$ for all $i \neq j$; that is, $\boldsymbol{\Sigma}$ is a diagonal matrix.

Proof. Independence immediately implies uncorrelatedness between components. For the opposite direction, recall the MVN distribution has density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

If $\boldsymbol{\Sigma}$ is diagonal, then $(\mathbf{x}-\boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \sum_{i=1}^n \frac{(x_i-\mu_i)^2}{\sigma_i^2}$ and $\det \boldsymbol{\Sigma} = \prod_{i=1}^n \sigma_i^2$. Therefore, the joint $f_{\mathbf{X}}(\mathbf{x})$ factorizes into the product of $N(\mu_i, \sigma_i^2)$ densities, which implies independence.
Remark: This is a super useful technique in checking independence regarding normal distributions. Instead of writing out densities (or MGF's/CF's) and see if they factorize, just check that the covariances between pairwise components of a MVN RV are 0. □

Theorem 8.1.2. Now let $X_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Then, $\overline{X} \perp\!\!\!\perp X_i - \overline{X}$ for all i.

Corollary 8.1.1. $\overline{X} \perp\!\!\!\perp S^2$

Proof. First, it is easy to see that

$$\begin{pmatrix} \overline{X} \\ X_i - \overline{X} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\mathbf{1}^\intercal \\ e_i^\intercal - \frac{1}{n}\mathbf{1}^\intercal \end{pmatrix} \mathbf{X}$$

where $\mathbf{1}$ is a n-vector of 1's and e_i is the basis vector with ith component equal 1, and elsewhere 0. Therefore $(\overline{X}, X_i - \overline{X})$ is a linearly transformation of a MVN, hence it is MVN. Now we can just apply the covariance condition in Theorem 8.1.1. We have

$$\text{Cov}(\overline{X}, X_i - \overline{X}) = \frac{1}{n}\text{Var}(X_i) - \text{Var}(\overline{X}) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$$

Remark: Suppose we have *heteroskedasticity*: $X_i \sim N(\mu, \sigma_i^2)$. Is it still true that $\overline{X} \perp\!\!\!\perp S^2$? Is S^2 still a scaled-χ^2 distribution? Does $t = \frac{\overline{X}-\mu}{\sqrt{S^2/n}}$ still follow a t-distribution? □

8.2. Jacobians

Definition 8.2.1. Suppose $g : \mathbf{X} \to \mathbf{Y}$, 1-1 mapping in the support of $\mathbf{X}, \mathbf{Y}$. Then

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y}))|\det J_{\mathbf{Y}\to\mathbf{X}}|^{-1} = f_{\mathbf{X}}(g^{-1}(\mathbf{y}))|\det J_{\mathbf{X}\to\mathbf{Y}}|$$

where

$$J_{\mathbf{Y}\to\mathbf{X}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left(\frac{\partial y_i}{\partial x_j}\right)_{ij} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \end{pmatrix}$$

The $\mathbf{Y} \to \mathbf{X}$ (and vice-versa) subscripts on the Jacobians are to help remember the direction you take derivatives in; it is most useful to calculate $|\det J_{\mathbf{X}\to\mathbf{Y}}|$, so that the RHS is in terms of $\mathbf{y}$ and we do not need to perform the inverse factor.

- Another way to remember (chain rule):

$$\int_A f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = \int_{g^{-1}(A)} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \int_A f_{\mathbf{X}}(g^{-1}(\mathbf{y}))\left|\det\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| d\mathbf{y}$$

  If we ignore the det for a moment, the $d\mathbf{y}$ "cancel" out and gives the integration on the scale of $d\mathbf{x}$.

- Geometric perspective: ratio of the area of regions under transformed coordinates

$$\text{vol}(d\mathbf{y}) = \left|\det\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right| \text{vol}(d\mathbf{x})$$

Example 8.2.1. Consider

$$\begin{aligned} X_1 &= R\cos(\Theta_1) \\ X_2 &= R\sin(\Theta_1)\cos(\Theta_2) \\ &\vdots \\ X_{n-1} &= R\sin(\Theta_1)\cdots\sin(\Theta_{n-2})\cos(\Theta_{n-1}) \\ X_n &= R\sin(\Theta_1)\cdots\sin(\Theta_{n-1}) \end{aligned}$$

known as the *generalized spherical transformation* from $(X_1, \cdots, X_n) \to (R, \Theta_1, \cdots, \Theta_{n-1}) \in \mathbb{R}_{\geq 0} \times [0, \pi]^{n-2} \times [0, 2\pi]$. We'll notate this as $\mathbf{X} \to (R, \boldsymbol{\Theta})$. It is straightforward to show that

$$|\det J_{\mathbf{X}\to(R,\boldsymbol{\Theta})}| = r^{n-1}\sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2)\cdots\sin(\theta_{n-2})$$

Example 8.2.2. Suppose $\mathbf{X}$ has pdf $f_{\mathbf{X}}(\mathbf{x})$ that satisfies, under re-parametrization, $f_{\mathbf{X}}(\mathbf{x}) = h(r)$. Determine the pdf of $R = \sqrt{X_1^2 + \cdots X_n^2}$.

Proof. First, the joint density is

$$f_{R,\boldsymbol{\Theta}}(r, \boldsymbol{\theta}) = f_{\mathbf{X}}(\mathbf{x})|\det J_{\mathbf{X}\to(R,\boldsymbol{\Theta})}|$$

$$= h(r)r^{n-1}\sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2)\cdots\sin(\theta_{n-2})$$

Then marginally,

$$\begin{aligned} f_R(r) &= \int_{[0,\theta]^{n-2}\times[0,2\pi]} f_{R,\boldsymbol{\Theta}}(r,\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= h(r)r^{n-1}\int_{[0,\theta]^{n-2}\times[0,2\pi]} \sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2)\cdots\sin(\theta_{n-2})d\boldsymbol{\theta} \\ &= h(r)r^{n-1}\frac{(2\pi)^{n/2}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} \end{aligned}$$

One example of $f_{\mathbf{X}}(\mathbf{x})$ satisfies the $h(r)$ condition is $X_1,\cdots,X_n \overset{\text{iid}}{\sim} N(0,1)$. Then,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} = \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}r^2} = h(r)$$

Hence

$$f_R(r) = h(r)r^{n-1}\frac{(2\pi)^{n/2}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} = \frac{r^{n-1}e^{-\frac{1}{2}r^2}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}}$$

which is the density of the χ-distribution (not χ^2!). □

Example 8.2.3. Let $X_1,\cdots,X_n \overset{\text{iid}}{\sim} f_X$. Consider the transformation g that performs $g(X_1,\cdots,X_n) = (X_{(1)},\cdots,X_{(n)})$. Then, the joint distribution of the order statistics is

$$f_{\mathbf{X}_{(n)}}(\mathbf{x}) = n!f_{\mathbf{X}}(\mathbf{x})\mathbb{I}(x_1 \le x_2 \le \cdots \le x_n)$$

In some sense, we can view $n!$ as the Jacobian factor $|\det J_{\mathbf{X}\to\mathbf{X}_{(n)}}|$, although deriving it analytically with the partial derivatives definition would be fruitless. Here are two heuristic proofs:

- Because X_i are iid, the distribution of all possible permutations of $\mathbf{X}_{(n)}$ is the distribution of $\mathbf{X}$. Hence, we need to multiply by the number of permutations, which is $n!$.
- Recall the geometric interpretation of Jacobians: $\text{vol}(d\mathbf{x}) = \left|\det J_{\mathbf{X}\to\mathbf{X}_{(n)}}\right| \text{vol}(d\mathbf{x}_{(n)})$. On the LHS, suppose $\text{vol}(d\mathbf{x}) \doteq \varepsilon$. On the RHS, $\text{vol}(d\mathbf{x}_{(n)})$ is approximately ε times a scale factor to account for the fact that $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$; that scale factor is the volume of the region $\{(x_1,\cdots,x_n) : x_1 \le x_2 \le \cdots \le x_n\}$, which is $\frac{1}{n!}$; hence $\left|\det J_{\mathbf{X}\to\mathbf{X}_{(n)}}\right| = n!$.

8.3. Additional problems

1. The sets

$$\mathcal{R}_x = \{(x_1, \cdots, x_n) : 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1\}$$
$$\mathcal{R}_y = \{(y_1, \cdots, y_n) : 0 \leq y_i \leq 1, \sum_{i=1}^{n} y_i \leq 1\}$$

are related through the transformation

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Note that $\mathcal{R}_y$ is the region associated with an n-dimensional unit simplex (look back to Example 3.3.2 from Lab 3). Show that $\text{vol}(\mathcal{R}_x) = \text{vol}(\mathcal{R}_y) = \frac{1}{n!}$.

2. We say $\mathbf{X}$ follows a uniform distribution *within* the unit n-sphere if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{V} \cdot \mathbb{I}(x_1^2 + \cdots + x_n^2 \leq 1)$$

for some constant V such that $\int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1$. Use generalized spherical coordinates to show that $V = \frac{2\pi^{n/2}}{n\Gamma(\frac{n}{2})} = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}$. Geometrically interpret V.

3. Similarly, we say that $\mathbf{X}$ follows a uniform distribution *on* the unit n-sphere if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{S} \cdot \mathbb{I}(x_1^2 + \cdots + x_n^2 = 1)$$

for some constant S such that $\int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1$. Show that $S = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}$. Geometrically interpret S.

4. Let $\mathbf{X} \sim N(\mathbf{0}, \mathcal{I}_n)$, and define

$$R = \|\mathbf{X}\| = \sqrt{X_1^2 + \cdots X_n^2} \qquad \text{and} \qquad \mathbf{U} = \frac{\mathbf{X}}{R} = \begin{pmatrix} \cos(\Theta_1) \\ \sin(\Theta_1)\cos(\Theta_2) \\ \vdots \\ \sin(\Theta_1)\cdots\sin(\Theta_{n-2})\cos(\Theta_{n-1}) \\ \sin(\Theta_1)\cdots\sin(\Theta_{n-1}) \end{pmatrix}$$

(a) Use generalized spherical coordinates to show the joint distribution for $(R, \boldsymbol{\Theta})$ factorizes: $f_{R,\boldsymbol{\Theta}}(r, \boldsymbol{\theta}) = f_R(r) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. Since $\mathbf{U}$ is only a function of $\boldsymbol{\Theta}$, conclude that $R \perp\!\!\!\perp \mathbf{U}$.

(b) Show that the density of $\mathbf{U}$ in terms of $\boldsymbol{\Theta}$ is

$$f_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{U}}(\mathbf{u}(\boldsymbol{\theta})) = \frac{\Gamma(\frac{n}{2})}{2\pi^{n/2}} \frac{1}{\sin(\theta_1)\sin(\theta_2)\cdots\sin(\theta_{p-2})|\sin(\theta_{p-1})|}$$