Section 4: Vector random variables

ARE 210

September 19, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section4" folder.

1 Definitions

- **Cauchy-Schwarz inequality**: $\mathbf{E}[XY]^2 \leq \mathbf{E}[X^2]\mathbf{E}[Y^2]$
  - $\mathbf{E}[XY]^2 = \mathbf{E}[X^2]\mathbf{E}[Y^2] \Leftrightarrow X = cY$ almost surely
- **Moment inequality**: $|\mathbf{E}[X^m]| < \infty, k < m \Rightarrow |\mathbf{E}[X^k]| < \infty$ ($\Leftarrow$ Jensen's)
- **Holder's inequality**: $p, q > 0$, $\frac{1}{p} + \frac{1}{q} = 1 \Rightarrow |\mathbf{E}[XY]| \leq \mathbf{E}[|XY|] \leq \mathbf{E}[|X|^p]^{1/p}\mathbf{E}[|X|^q]^{1/q}$
  - Cauchy-Schwarz inequality when $p = q = 2$
- **Minkowski's inequality**: $\mathbf{E}[|X+Y|^p]^{1/p} \leq \mathbf{E}[|X|^p]^{1/p} + \mathbf{E}[|Y|^p]^{1/p}$
- Joint distributions
  - $F(x) = P[X_1 \leq x_1, \ldots, X_n \leq x_n]$ is the **joint CDF**
  - $F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(s) ds_n \cdots ds_1$ implicitly defines the **joint PDF**, f
  - $F_j(x_j) = P[X_j \leq x_j] = \lim\limits_{i \to \infty, i \neq j} F(x)$ is the **marginal distribution**
  - The **conditional distribution** of a random variable Y given X is a measure $\mu_{Y|X}(\cdot|x)$ satisfying, $\forall\ A, B,\ P[X \in B, Y \in A] = \int_B \mu_{Y|X}(A, x) F_{X,Y}(dx)$, where $F_{X,Y}$ is the joint CDF of the random variable (X, Y)
    - Simplifies to Bayes' Rule + Law of Total Probability when X is discrete
    - Suppose (X, Y) has a nicely behaved density $f_{X,Y}$. Let f_X be the marginal density. Then, Y conditional on $X = x$ has **conditional density** $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
    - The **conditional expectation** of a random variable $Y|X$ is $\mathbf{E}[Y|X = x] = \int y F_{Y|X}[dy|x] \equiv \int y f_{Y|X}(y|x) dy$, where $F_{Y|X}[\cdot, x]$ is the conditional probability measure for $Y|X = x$ and $f_{Y|X}(\cdot|x)$ is the conditional density
  - Multivariate normal (MVN)
    - $X_{d\times 1}$ is MVN $\Leftrightarrow A_{m\times d}X_{d\times 1}$ is MVN $\forall\ A_{m\times d}$
    - X MVN $\Leftrightarrow \Psi_X(r) = \exp[ir^{\mathrm{T}}\mu - r^{\mathrm{T}}Vr/2]$, where Ψ_X is the characteristic function of X, $\mu = \mathbf{E}[X]$, and $V = V[X]$
    - X MVN $\Rightarrow (X_1 \perp X_2 \Leftrightarrow \mathrm{Cov}[X_1, X_2] = 0)$

- Matrices
  - λ is an **eigenvalue** of M if $\exists v \neq 0$ such that $Mv = \lambda v$
  - M **PD** (**PSD**) $\Leftrightarrow$ all eigenvalues $>$ $(\geq)$ 0
  - M PD $\Leftrightarrow$ M PSD, M^{-1} exists
  - Let $V_{d\times d}$ be a PD matrix, $\mu_{d\times 1}$ a vector, then the density of a MVN random variable with variance V and mean μ is

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{\det[V]}} \exp\left[-\frac{1}{2}(x-\mu)^{\mathrm{T}} V^{-1} (x-\mu)\right]$$

- The **empirical cdf** is $F_n(x,\omega) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}[X_i(\omega) \leq x]$
  - Equivalent measure $\mu_n(A,\omega) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}[X_i(\omega) \in A]$

2 Some helpful tips

- Equality of marginal distributions $\not\Rightarrow$ equality of joint distribution
- Conditional distributions are probability measures
- We said M PD $\Rightarrow$ $M = KDK^{\mathrm{T}}$ where D is a diagonal matrix of eigenvalues and K is orthonormal ($KK^{\mathrm{T}} = \mathbf{I}$). Basically, in the expression $Mv = KDK^{\mathrm{T}}v$ think of K and K^{T} as changing the basis of the vector v to the basis of eigenvectors, M is simply scaling each eigenvector comprising $K^{\mathrm{T}}v$ by its eigenvalue, after which K returns it to the original basis.

3 Practice questions

1) **MVN characteristic function**: Derive the characteristic function for a multivariate normal random variable with mean $\mathbf{0}_{n\times 1}$ and variance $\mathbf{I}_{n\times n}$

2) **Bayesian updating**: Suppose I have a prior belief that some population parameter $\theta \sim N(0,1)$. Suppose I observe an unbiased signal of θ, t, with normally distributed noise with variance σ_t^2. What are $f(t|\theta)$ and $f(\theta)$? Derive $f(t)$ and $f(\theta|t)$. Interpret $f(\theta|t)$.