$\frac{4}{8}$ $\frac{5}{2}$ $\boxed{\frac{6}{9}}$ $\frac{7}{6}$ $\frac{8}{4}$

August 8, 2019

TOWER PROPERTY OF CONDITIONAL EXPECTATION:

A fair coin is tossed until two tails occur successively. Let N be the total number of tosses required to terminate the experiment. Compute $\mathbb{E}[N]$.

The essence of the tower property is the following:

$$\underbrace{\mathbb{E}}_{\text{w.r.t. } X}\Big[\underbrace{\mathbb{E}[N|X]}_{\text{w.r.t. } N,\text{ given } X}\Big]$$

$$\mathbb{E}[N] = \mathbb{E}\big[\mathbb{E}[N|X]\big]$$

w.r.t. N, given X

w.r.t. X

Obviously, X must be wisely chosen to be helpful.

$$\text{Let } X = \begin{cases} 1 & \text{if the first toss results in tails} \\ 0 & \text{if the first toss results in heads} \end{cases}$$

indicator r.v.
dummy r.v.

$$\mathbb{E}[N] = \mathbb{E}\big[\mathbb{E}[N|X]\big] \quad \text{(work from outside, in)}$$

$$= \mathbb{E}[N|X=1]\,\mathbb{P}(X=1) + \mathbb{E}[N|X=0]\,\mathbb{P}(X=0)$$

$$= \Big(\big(\tfrac{1}{2}\big)\underbrace{\big(2+\mathbb{E}[N]\big)}_{\text{heads on second flip}} + \big(\tfrac{1}{2}\big)\underbrace{(2)}_{\text{tails on second flip}}\Big)\cdot\tfrac{1}{2} + \big(1+\mathbb{E}[N]\big)\cdot\tfrac{1}{2}$$

heads ~~tails~~ on second flip

tails ~~heads~~ on second flip

$$\Rightarrow \mathbb{E}[N] = \big(1 + \cancel{\mathbb{E}}\ \tfrac{1}{2}\mathbb{E}[N] + 1\big)\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2}\mathbb{E}[N]$$

$$\Rightarrow \mathbb{E}[N] = 1 + \tfrac{1}{4}\mathbb{E}[N] + \tfrac{1}{2} + \tfrac{1}{2}\mathbb{E}[N] \Rightarrow \tfrac{1}{4}\mathbb{E}[N] = \tfrac{3}{2} \Rightarrow \mathbb{E}[N] = 6$$

DEF: If $\vec{X} \sim N(\vec{\mu}, \Sigma)$, we say that $\vec{X}$ is a multivariate Gaussian random vector with mean vector $\vec{\mu}$ and variance-covariance matrix Σ.

NOTE: If $\vec{x}$ is k-dimensional, the $\vec{\mu} \in \mathbb{R}^k$ and $\Sigma \in \mathbb{R}^{k \times k}$. Moreover, Σ must be positive semi-definite.

RECALL: Σ is positive semi-definite iff $\forall \vec{x} \in \mathbb{R}^k$,

$$\vec{x}^T \Sigma \vec{x} \geq 0.$$

RECALL: If all eigenvalues are ≥ 0, then Σ is positive ~~&~~ semi-definite.

QN: What is the density function for $\vec{X}$?

ANS: $f(x_1, \ldots, x_k) = f(\vec{x}) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right\}$

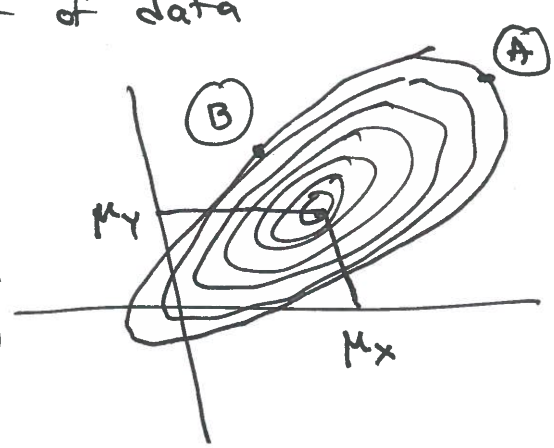
FUN FACT: Call $d(\vec{x}, \vec{y}) = (\vec{x}-\vec{y})^T \Sigma^{-1} (\vec{x}-\vec{y})$.

this is the variance-covariance matrix of some underlying set of data

This distance is called Mahalanobis distance.

$\Rightarrow$ key role in discriminant analysis

$\Rightarrow$ key role in k-means clustering



A is farther than B from (μ_x, μ_y).

A and B are equidistant from (μ_x, μ_y) under Mahalanobis.

FACTS ABOUT MULTIVARIATE GAUSSIAN RANDOM VECTORS

Suppose that $\vec{X} = (X_1, \ldots, X_k)$ with $\vec{X} \sim N(\vec{\mu}, \Sigma)$.

① Every linear combination of components of $\vec{X}$ is again Gaussian. In other words,

$$a_1 X_1 + a_2 X_2 + \ldots + a_k X_k = \vec{a}^T \vec{X} \leftarrow \text{this is Gaussian}$$

$$\Rightarrow \mathbb{E}\left[\vec{a}^T \vec{X}\right] = \vec{a}^T . \mathbb{E}\left[\vec{X}\right] = \vec{a}^T . \mu$$

$$\Rightarrow \operatorname{Var}(\vec{X}) = \Sigma$$

↖ by definition, i.e., this is how we will extend the notion of variance to random vectors

$$\Rightarrow \operatorname{Var}(\vec{a}^T . \vec{X}) \cancel{= \vec{a}^T \vec{a}\, \Sigma} = \vec{a}^T \Sigma\, \vec{a}$$

② For any such $\vec{X}$, there exists a matrix A so that

$$\vec{X} = A . \vec{Z} + \vec{\mu}$$

where $\vec{Z}$ is a k-dimensional vector of independent standard normal random variables.

Clearly: A is related to a "square root" of Σ, which Σ always has because (1) it is positive semi-definite and ② it is symmetric.

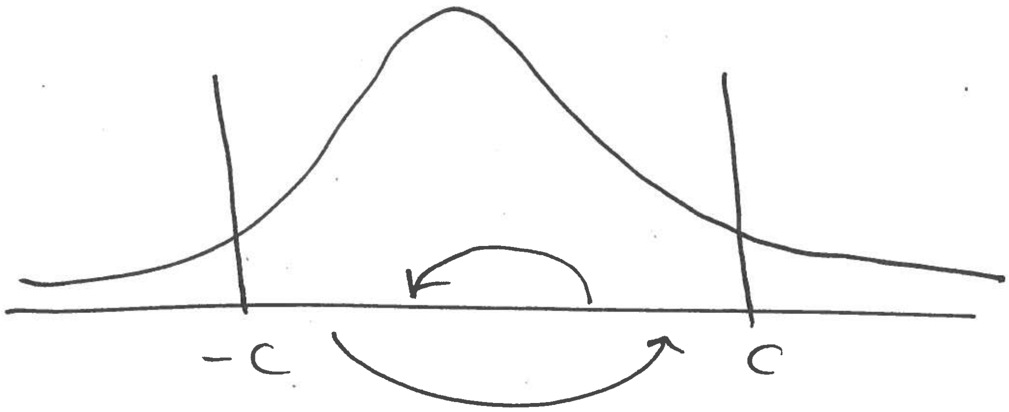
③ For any k, the level curves (or level sets) of the density $f(\vec{x})$ are ellipsoids.

④ If X_i and X_j are such that $\rho(X_i, X_j) = 0$, then X_i and X_j are independent.

⑤ Two random variables that are normally distributed are not necessarily multivariate Gaussian.

EX: Let $X \sim N(0,1)$.

Define $Y = \begin{cases} X & \text{if } |X| > c \\ -X & \text{if } |X| \le c \end{cases}$



Clearly $Y \sim N(0,1)$.

Unfortunately, (X,Y) is NOT bivariate normal.

Qn: Why can't X and Y be jointly normal?

$\Rightarrow$ b/c $\mathrm{CORR}(X,Y) = +1$ when $|X| > c$

$\mathrm{CORR}(X,Y) = -1$ when $|X| \le c$

(A multivariate Gaussian has the same covariances, or same correlations, throughout the support set of the density.)

Asymptotic Distribution of an MLE

Suppose that $X_1, \ldots, X_n$ is a random sample from some density $f(x; \vec{\theta})$. Let $\hat{\vec{\theta}}$ be an MLE vector for $\vec{\theta}$. Under a (fairly easy to satisfy) set of conditions mentioned on the homework,

$$\sqrt{n}\left(\hat{\vec{\theta}} - \vec{\theta}\right) \sim N\left(0, I^{-1}\right)$$

inverse of the Fisher transformation; see more on the homework

In the case when $\vec{\theta} = \theta$ is one-dimensional,

$$I = \mathbb{E}\left[\underbrace{\frac{\partial}{\partial\theta} \log f(X;\theta) \cdot \frac{\partial}{\partial\theta} \log\left(\frac{\partial}{\partial x} f(x;\theta)\Big|_{x=X}\right)}_{\text{Cramer-Rao Inequality -or- Cramer-Rao Lower Bound}}\right]$$

Cramer-Rao Inequality -or- Cramer-Rao Lower Bound

BIG PICTURE POINT: whenever you are estimating a k-dimensional parameter vector $\vec{\theta}$ (e.g., in time series class, linear regression class), that vector of MLEs $\hat{\vec{\theta}}$ is asymptotically multivariate Gaussian.

↪ its variance-covariance matrix will be useful

↪ door is opened to normality-based CIs and hypothesis tests