# Information Measure of Dependence: Some Virtues and a Caveat

**Ehsan Soofi**
Lubar School, University of Wisconsin-Milwaukee

Presentation at the IPM Isfahan

Coauthors:
Nader Ebrahimi, Northern Illinois University
Nima Jalali, Lubar School of Business

December 25, 2012

# Outline

- Notions of dependence, association, and predictability
- Information approach to dependence
  - Scale of predictability
  - The mutual information
    - Utility of dependence
    - Dependence information index (absolutely continuous distributions)
- Failure of traditional measures to capture dependence
- Location-scale family
  - Gaussian, Student-$t$, Elliptical
- Dependence between sum and summands
  - Regression (normal and beyond)
  - Stochastic processes
  - Measurement error

# Outline

- Notions of dependence, association, and predictability
- Information approach to dependence
  - Scale of predictability
  - The mutual information
    - Utility of dependence
    - Dependence information index (absolutely continuous distributions)
- Failure of traditional measures to capture dependence
- Location-scale family
  - Gaussian, Student-$t$, Elliptical
- Dependence between sum and summands
  - Regression (normal and beyond)
  - Stochastic processes
  - Measurement error
- Information index for singular models (if time allows)
  - Marshall-Olkin family
  - Gaver-Lewis family
  - Test of sharp hypothesis

# Dependence

- $X_1$ and $X_2$ two random variables
  - $(X_1, X_2)$, random vector with a bivariate $F$ with pdf $f$
  - Marginal distributions $F_i$, with pdf $f_i, i = 1, 2$
  - Conditional distributions with pdf's $f_{i|j}(x_i|x_j) = \dfrac{f(x_1, x_2)}{f_j(x_j)}$

# Dependence

- $X_1$ and $X_2$ two random variables
  - $(X_1, X_2)$, random vector with a bivariate $F$ with pdf $f$
  - Marginal distributions $F_i$, with pdf $f_i, i = 1, 2$
  - Conditional distributions with pdf's $f_{i|j}(x_i|x_j) = \dfrac{f(x_1, x_2)}{f_j(x_j)}$
- Independence is a stochastic notion

$$F(x_1, x_2) = F_1(x_1)F_2(x_2) \quad \text{for all } (x_1, x_2) \in \Re^2$$

**A sharp state**

# Dependence

- $X_1$ and $X_2$ two random variables
  - $(X_1, X_2)$, random vector with a bivariate $F$ with pdf $f$
  - Marginal distributions $F_i$, with pdf $f_i, i = 1, 2$
  - Conditional distributions with pdf's $f_{i|j}(x_i|x_j) = \dfrac{f(x_1, x_2)}{f_j(x_j)}$
- Independence is a stochastic notion

$$F(x_1, x_2) = F_1(x_1)F_2(x_2) \quad \text{for all } (x_1, x_2) \in \Re^2$$

**A sharp state**

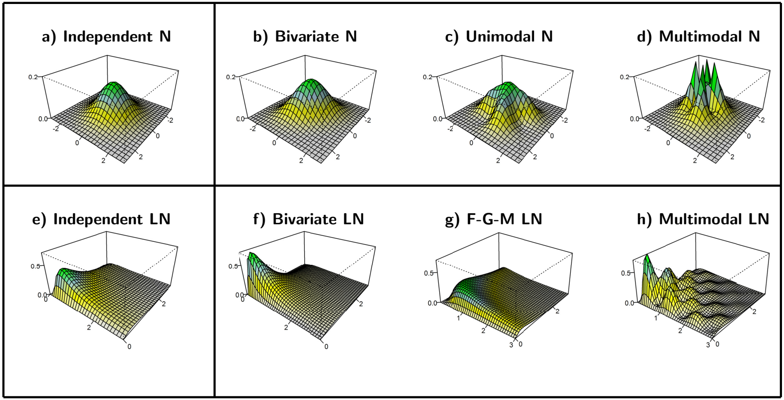
- Dependence is a negation of the independence

$$F(x_1, x_2) \neq F_1(x_1)F_2(x_2) \quad \text{for some } (x_1, x_2) \in \Re^2$$

**A multifarious notion**

# Examples of Bivariate Distributions

**Independent models** **Dependent models**




**Normal**

**Log-normal**

# Association

- Association is a covariation

$$\mathrm{cov}[\phi_1(X_1), \phi_2(X_2)] \neq 0$$

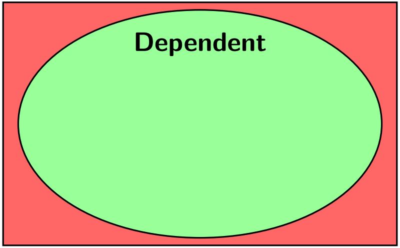
  - $\phi_i(\cdot),\ i = 1, 2$ monotone functions
  - Positive and negative $\mathrm{cov}[\phi_1(X_1), \phi_2(X_2)]$ are called positive and negative quadrant dependence in reliability
  - Correlation: linear association $\mathrm{cov}(X_1), X_2) \neq 0$

# Association

- Association is a covariation

$$\text{cov}[\phi_1(X_1), \phi_2(X_2)] \neq 0$$

  - $\phi_i(\cdot), \;\; i = 1, 2$ monotone functions
  - Positive and negative $\text{cov}[\phi_1(X_1), \phi_2(X_2)]$ are called positive and negative quadrant dependence in reliability
  - Correlation: linear association $\text{cov}(X_1), X_2) \neq 0$
- Diagram of relationships between two random variables

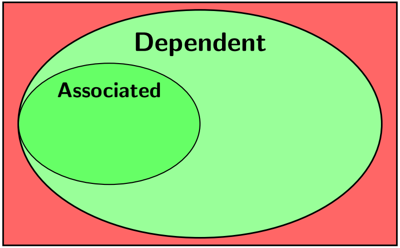


# Association

- Association is a covariation

$$\text{cov}[\phi_1(X_1), \phi_2(X_2)] \neq 0$$

  - $\phi_i(\cdot), \ i = 1, 2$ monotone functions
  - Positive and negative $\text{cov}[\phi_1(X_1), \phi_2(X_2)]$ are called positive and negative quadrant dependence in reliability
  - Correlation: linear association $\text{cov}(X_1), X_2) \neq 0$
- Diagram of relationships between two random variables

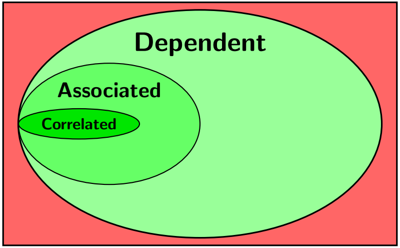


# Association

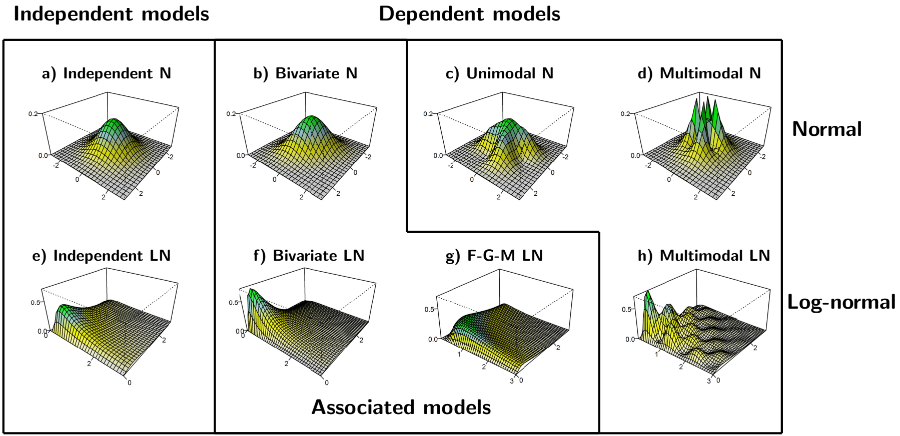
- Association is a covariation

$$\text{cov}[\phi_1(X_1), \phi_2(X_2)] \neq 0$$

  - $\phi_i(\cdot), \; i = 1, 2$ monotone functions
  - Positive and negative $\text{cov}[\phi_1(X_1), \phi_2(X_2)]$ are called positive and negative quadrant dependence in reliability
  - Correlation: linear association $\text{cov}(X_1), X_2) \neq 0$
- Diagram of relationships between two random variables



# Examples of Bivariate Distributions



# Some questions

- How much each dependent model deviates from its independent version?

# Some questions

- How much each dependent model deviates from its independent version?
- Which of the dependent models represents a weaker or a stronger dependence?

# Some questions

- How much each dependent model deviates from its independent version?
- Which of the dependent models represents a weaker or a stronger dependence?
- Which of these models would enable you to better predict one of the variables by using the other?

# Some questions

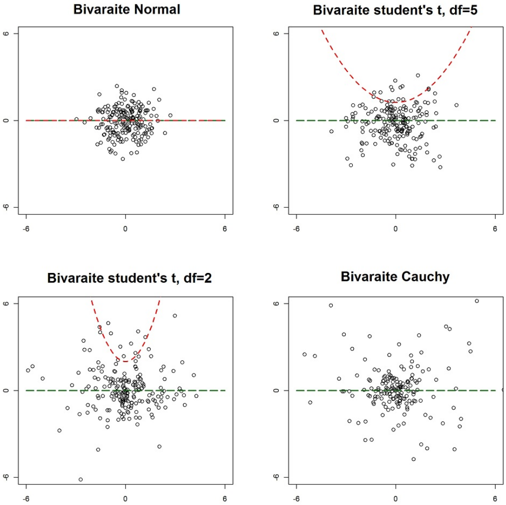
- How much each dependent model deviates from its independent version?
- Which of the dependent models represents a weaker or a stronger dependence?
- Which of these models would enable you to better predict one of the variables by using the other?
- Are more strongly associated models also more dependent?

# Some questions

- How much each dependent model deviates from its independent version?
- Which of the dependent models represents a weaker or a stronger dependence?
- Which of these models would enable you to better predict one of the variables by using the other?
- Are more strongly associated models also more dependent?
- The information notion of dependence answers these questions based on the departure of the joint distribution $F$ from the independent model $G = F_1 F_2$.

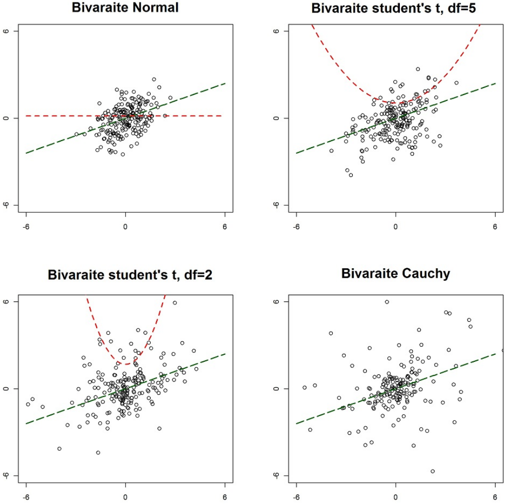
# Data From Four Unassociated Models

- Bivariate normal: Independent
- Bivariate $t$: Not independent



# Data From Four Equally Associated Models

- Regression $E(X_1|x_2) = .4x_2$
- Bivariate normal: Constant conditional variance
- Bivariate $t$: Quadratic conditional variance (not defined for Cauchy)



# Dependence & Predictability

- The independent state:
  - Complete absence of probabilistic information about the outcomes of each variable by the other
  - **Perfect unpredictability**

# Dependence & Predictability

- The independent state:
  - Complete absence of probabilistic information about the outcomes of each variable by the other
  - **Perfect unpredictability**
- Complete dependence (Lancaster 1963):

$$P[X_i = \phi_i(X_j)] = 1, \quad i \neq j$$

  - $\phi_i(\cdot)$, a measureable one-to-one function
  - **Perfect predictability** (Kimeldorf & Sampson 1978)

# Dependence & Predictability

- The independent state:
  - Complete absence of probabilistic information about the outcomes of each variable by the other
  - **Perfect unpredictability**
- Complete dependence (Lancaster 1963):

$$P[X_i = \phi_i(X_j)] = 1, \quad i \neq j$$

  - $\phi_i(\cdot)$, a measureable one-to-one function
  - **Perfect predictability** (Kimeldorf & Sampson 1978)
- Scale of predictability

**Perfect unpredictability** | **Perfect predictability**

**0** ——— **100%**

# Dependence & Predictability

- The independent state:
  - Complete absence of probabilistic information about the outcomes of each variable by the other
  - **Perfect unpredictability**
- Complete dependence (Lancaster 1963):

$$P[X_i = \phi_i(X_j)] = 1, \quad i \neq j$$

  - $\phi_i(\cdot)$, a measureable one-to-one function
  - **Perfect predictability** (Kimeldorf & Sampson 1978)
- Scale of predictability

**Perfect unpredictability** | **Perfect predictability**

**0** ——————————————— **100%**

**Independence**

**If and only if for all $(x_i, x_j)$,**

$$F(x_1, x_2) = F_1(x_1)F_2(x_2)$$

$$F(x_i|x_j) = F_i(x_i)$$

# Dependence & Predictability

- The independent state:
  - Complete absence of probabilistic information about the outcomes of each variable by the other
  - **Perfect unpredictability**
- Complete dependence (Lancaster 1963):

$$P[X_i = \phi_i(X_j)] = 1, \quad i \neq j$$

  - $\phi_i(\cdot)$, a measureable one-to-one function
  - **Perfect predictability** (Kimeldorf & Sampson 1978)
- Scale of predictability

**Perfect unpredictability** | **Perfect predictability**

**0** — **100%**

**Independence**

**If and only if for all $(x_i, x_j)$,**

$$\mathbf{F(x_1, x_2) = F_1(x_1)F_2(x_2)}$$
$$\mathbf{F(x_i|x_j) = F_i(x_i)}$$

**Complete dependence**

**For a one-to-one function $\phi_i(\cdot)$**

$$\mathbf{P[X_i = \phi_i(X_j)] = 1, i \neq j}$$

# Information Approach to Dependence

- The information notion of dependence compares $F$ with the independent model $G = F_1 F_2$
- The strength of dependence is measured by a divergence function between $\mathcal{D}(F : F_1 F_2) \geq 0$
  - the equality holds if and only if $f(x_1, x_2) = f_1(x_1) f_2(x_2)$ for almost all $(x_1, x_2) \in \Re^2$

# Information Approach to Dependence

- The information notion of dependence compares $F$ with the independent model $G = F_1 F_2$
- The strength of dependence is measured by a divergence function between $\mathcal{D}(F : F_1 F_2) \geq 0$
  - the equality holds if and only if $f(x_1, x_2) = f_1(x_1) f_2(x_2)$ for almost all $(x_1, x_2) \in \Re^2$
- Scale of $\mathcal{D}(F : F_1 F_2)$

**Perfect unpredictability** — **Perfect predictability**

**0** — $\mathcal{D}(\mathbf{F} : \mathbf{F}_1 \mathbf{F}_2)$ — **100%**

**Independence**

**If and only if for all $(x_i, x_j)$,**

$$F(x_1, x_2) = F_1(x_1) F_2(x_2)$$
$$F(x_i | x_j) = F_i(x_i)$$

**Complete dependence**

**For a one-to-one function $\phi_i(\cdot)$**

$$P[X_i = \phi_i(X_j)] = 1, i \neq j$$

**(Lancaster 1963)**
**(Kimeldorf & Sampson 1978)**

# Kullback-Leibler (KL) Information & Shannon Entropy

- The most well-known and widely-used divergence and uncertainty functions
- The KL information divergence

$$K(F:G) = \int_{\mathcal{S}} f(x) \log \frac{f(x)}{g(x)} dx$$

provided that the integral is finite

  - $\mathcal{S}$ is the support of $F$, provided that the integral is finite
  - $F$ must be absolutely continuous with respect to $G$, denoted $F \ll G$
  - It is also known as cross-entropy and relative entropy
  - Shannon entropy (Shannon 1948)

$$H(X) = H(F) = -\int_{\mathcal{S}} f(x) \log f(x) dx$$

    - $\text{Var}(X) < \infty \Rightarrow H(X) < \infty$, converse does not hold

# The mutual information

- The mutual information of the bivariate distribution $F$:

$$
\begin{aligned}
M(F) &\equiv M(X_1, X_2) \\
&= K(F : F_1 F_2) \\
&= E_{x_j}\Big\{H[F_i(x_i)] - H[F_{i|j}(x_i|X_j)]\Big\}, \quad j \neq i = 1, 2 \\
&= E_{x_j}\Big\{K[F_{i|j}(x_i|X_j) : F_i(x_i)]\Big\}, \quad j \neq i = 1, 2,
\end{aligned}
$$

provided that $F \ll F_1 F_2$

# The mutual information

- The mutual information of the bivariate distribution $F$:

$$\begin{aligned} M(F) &\equiv M(X_1, X_2) \\ &= K(F : F_1 F_2) \\ &= E_{x_j}\Big\{H[F_i(x_i)] - H[F_{i|j}(x_i|X_j)]\Big\}, \quad j \neq i = 1, 2 \\ &= E_{x_j}\Big\{K[F_{i|j}(x_i|X_j) : F_i(x_i)]\Big\}, \quad j \neq i = 1, 2, \end{aligned}$$

provided that $F \ll F_1 F_2$

  - $M(X_1, X_2) \geq 0$, equality holds $\Leftrightarrow$ $X_1$ and $X_2$ are independent.

# The mutual information

- The mutual information of the bivariate distribution $F$:

$$\begin{aligned} M(F) &\equiv M(X_1, X_2) \\ &= K(F : F_1 F_2) \\ &= E_{x_j}\Big\{H[F_i(x_i)] - H[F_{i|j}(x_i|X_j)]\Big\}, \quad j \neq i = 1, 2 \\ &= E_{x_j}\Big\{K[F_{i|j}(x_i|X_j) : F_i(x_i)]\Big\}, \quad j \neq i = 1, 2, \end{aligned}$$

provided that $F \ll F_1 F_2$

  - $M(X_1, X_2) \geq 0$, equality holds $\Leftrightarrow$ $X_1$ and $X_2$ are independent.
  - The absolute continuity requires $P[X_1 = \phi_1(X_2)] = 0$ and $P[X_2 = \phi_2(X_1)] = 0$
    - Inapplicable to singular distributions

# The mutual information

- The mutual information of the bivariate distribution $F$:

$$\begin{aligned} M(F) &\equiv M(X_1, X_2) \\ &= K(F : F_1 F_2) \\ &= E_{x_j}\left\{ H[F_i(x_i)] - H[F_{i|j}(x_i|X_j)] \right\}, \quad j \neq i = 1, 2 \\ &= E_{x_j}\left\{ K[F_{i|j}(x_i|X_j) : F_i(x_i)] \right\}, \quad j \neq i = 1, 2, \end{aligned}$$

provided that $F \ll F_1 F_2$

  - $M(X_1, X_2) \geq 0$, equality holds $\Leftrightarrow$ $X_1$ and $X_2$ are independent.
  - The absolute continuity requires $P[X_1 = \phi_1(X_2)] = 0$ and $P[X_2 = \phi_2(X_1)] = 0$
    - Inapplicable to singular distributions
  - Lindley's (1956) Bayesian measure of sample information about a parameter $M(X, \Theta)$
    - The expected utility interpretation (Bernardo 1979)

# Utility of dependence

- The predictability of outcomes of $X_i$ without using $X_j, j \neq i = 1, 2$, depends solely on the concentration of its marginal distribution. $H(X_i)$ measures this uncertainty.

# Utility of dependence

- The predictability of outcomes of $X_i$ without using $X_j, j \neq i = 1, 2$, depends solely on the concentration of its marginal distribution. $H(X_i)$ measures this uncertainty.
- Given an outcome $x_j$ of $X_j$, the predictability of outcomes of $X_i$ depends on, the concentration of the conditional distribution of $X_i$ given $X_j = x_j$, measured by $H(F_{i|j})$.

# Utility of dependence

- The predictability of outcomes of $X_i$ without using $X_j, j \neq i = 1, 2$, depends solely on the concentration of its marginal distribution. $H(X_i)$ measures this uncertainty.
- Given an outcome $x_j$ of $X_j$, the predictability of outcomes of $X_i$ depends on, the concentration of the conditional distribution of $X_i$ given $X_j = x_j$, measured by $H(F_{i|j})$.
- The *expected utility* of $X_j$ for prediction of $X_i$ is given by the representation

$$M(X_1, X_2) = E_{x_j}\left\{H[F_i(x_i)] - H[F_{i|j}(x_i|X_j)]\right\}, \quad j \neq i = 1, 2$$

  - The bracketed quantity is known as the observed information provided by $x_j$ for predicting $X_i$.

# Utility of dependence

- The predictability of outcomes of $X_i$ without using $X_j, j \neq i = 1, 2$, depends solely on the concentration of its marginal distribution. $H(X_i)$ measures this uncertainty.
- Given an outcome $x_j$ of $X_j$, the predictability of outcomes of $X_i$ depends on, the concentration of the conditional distribution of $X_i$ given $X_j = x_j$, measured by $H(F_{i|j})$.
- The *expected utility* of $X_j$ for prediction of $X_i$ is given by the representation

$$M(X_1, X_2) = E_{x_j}\left\{H[F_i(x_i)] - H[F_{i|j}(x_i|X_j)]\right\}, \quad j \neq i = 1, 2$$

  - The bracketed quantity is known as the observed information provided by $x_j$ for predicting $X_i$.
- **When two variables are dependent, one is useful for predicting the other, irrespective of whether or not being associated**

# Other Divergence measures

- Among the known divergence measures and generalizations of Shannon entropy, only the KL information admits the expected utility representation
- The immediate generalizations are Rènyi measures

$$K_r(F:G) = \frac{1}{r-1}\log\int f^r(x)[g(x)]^{1-r}dx, \quad r \neq 1, \quad r > 0,$$
$$H_r(X) = \frac{1}{1-r}\log\int f^r(x)dx, \quad r \neq 1, \quad r > 0$$

  - $K_1(F:G) = K(F:G), H_1(X) = H(X)$
  - Example: Bivariate normal distribution with correlation $\rho$:

$$E_{x_j}\left\{H_r[F_i(x_i)] - H_r[F_{i|j}(x_i|X_j)]\right\} = M(X_1, X_2) = -\frac{1}{2}\log(1-\rho^2)$$
$$K_r(F:F_1,F_2) = M(X_1,X_2) + \frac{1}{2(1-r)}\log\left(1-(1-r)^2\rho^2\right)$$

Discrepancy with the independent normal is more $(r > 1)$ or less $(r < 1)$ than the utility

# Fourth Representation

- The unique additive property of Shannon entropy also gives the following representation

$$M(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2),$$

  - Shared or redundant information
  - The finiteness of the joint and marginal entropies are necessary. However, this is not sufficient
  - Particularly useful for calculating $M$ by entropy expressions

# Copula information

- $M(X_1, X_2)$ is invariant under 1-to-1 transformations of each $X_i$
- Copula of $F$
  - Let $U_i = F_i(X_i),\ i = 1, 2.$ Then

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \quad (u_1, u_2) \in [0, 1]^2$$

    (Sklar 1959)
  - A widely used approach for modeling dependence

# Copula information

- $M(X_1, X_2)$ is invariant under 1-to-1 transformations of each $X_i$
- Copula of $F$
  - Let $U_i = F_i(X_i),\ i = 1, 2.$ Then

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \quad (u_1, u_2) \in [0, 1]^2$$

    (Sklar 1959)
  - A widely used approach for modeling dependence
- Copula information

$$M(F) = M(C) = K(C : C_0) = -H(C) = I(C) \geq 0$$

  - $C_0$ denotes the product copula $C_0(u_1, u_2) = u_1 u_2$.
  - $I(C)$ is referred to as the information measure of the distribution (Lindley 1956, Zellner 1971), here the copula.

# Dependence Information Index

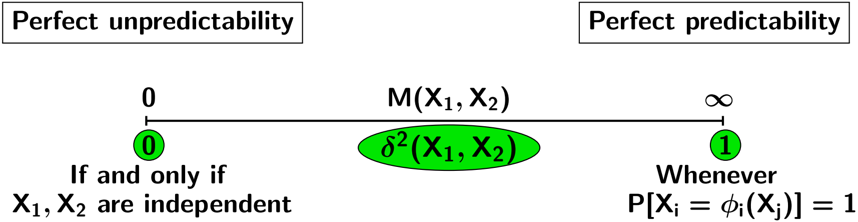
**Perfect unpredictability** | **Perfect predictability**

$0$ — $M(X_1, X_2)$ — $\infty$

**If and only if**
**$X_1, X_2$ are independent**

**Whenever**
**$P[X_i = \phi_i(X_j)] = 1$**

# Dependence Information Index




- For the *absolutely continuous* distributions:

$$\delta^2(F) = \delta^2(X_1, X_2) = 1 - e^{-2M(X_1,X_2)}$$

- Entropy reduction

$$\delta^2(X_1, X_2) = 1 - \frac{\exp\{E_{x_j}[H(X_i|x_j)]\}^2}{\exp\{H(X_j)\}^2} = 1 - \frac{\exp\{H(X_1, X_2)\}^2}{\exp\{H(X_1) + H(X_2)\}^2}$$

- Copula representation

$$\delta^2(F) = \delta^2(C) = 1 - e^{-2I(C)}$$

# Moment-based Indices

- Two popular indices
  - Pearson correlation coefficient $\rho_p$:
    - Subscript is for distinction between the correlation coefficient and a model parameter $\rho$
    - $E(X_i X_j) < \infty, \; i, j = 1, 2$
    - Invariant under linear transformations (up to the sign)
  - The fraction of expected variance reduction due to regression, also known as the correlation fraction:

$$\eta^2_{i|j} = 1 - \frac{E_{x_j}[\mathrm{Var}(X_i|x_j)]}{\mathrm{Var}(X_i)} \geq 0, \quad j \neq i = 1, 2,$$

    - $\mathrm{Var}(X_i), \mathrm{Var}(X_i|x_j) < \infty, \; i, j = 1, 2$
    - Invariant under linear transformations

# Information Approach: Ways beyond the bivariate normal

- For the bivariate normal distribution
$$\delta^2(F) = \eta^2(F) = \rho_p^2(F) = \rho^2$$
- Information approach
  - **Linear relationship** generalizes to **any functional relationship**.

# Information Approach: Ways beyond the bivariate normal

- For the bivariate normal distribution
$$\delta^2(F) = \eta^2(F) = \rho_p^2(F) = \rho^2$$
- Information approach
  - **Linear relationship** generalizes to **any functional relationship**.
  - **Variance** generalizes to an **uncertainty function**
    (a) Concave function of $f$ that measures the concentration, $H(F) \leq H(\text{uniform})$
    (b) The variance does not always satisfy this condition

# Information Approach: Ways beyond the bivariate normal

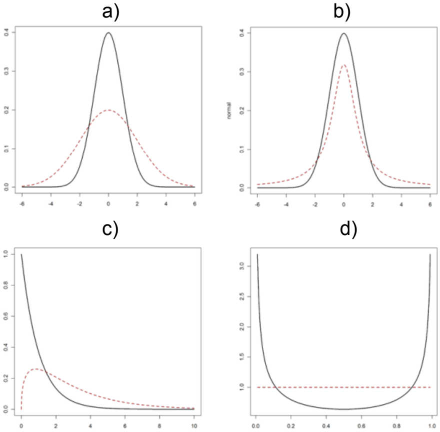
- For the bivariate normal distribution
$$\delta^2(F) = \eta^2(F) = \rho_p^2(F) = \rho^2$$
- Information approach
  - **Linear relationship** generalizes to **any functional relationship**.
  - **Variance** generalizes to an **uncertainty function**
    (a) Concave function of $f$ that measures the concentration, $H(F) \le H(\text{uniform})$
    (b) The variance does not always satisfy this condition
  - **The departure from independence** is measured formally by a **divergence function** between two probability distributions $\mathcal{D}(P : Q) \ge 0$, where $\mathcal{D} = 0$ if and only if the distributions are identical $dP(x) = dQ(x)$

# Information Approach: Ways beyond the bivariate normal

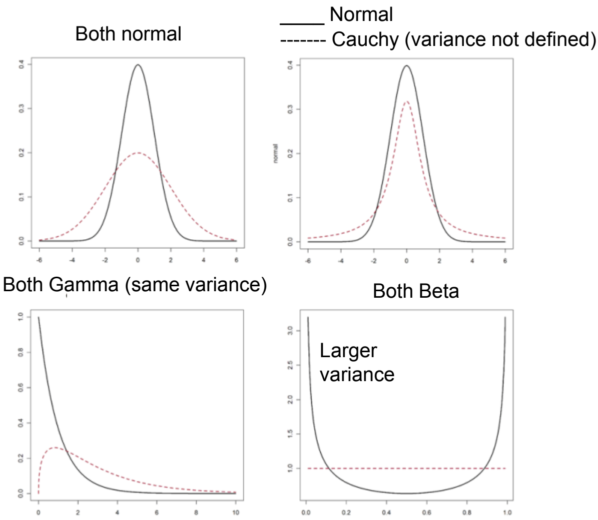
- For the bivariate normal distribution
$$\delta^2(F) = \eta^2(F) = \rho_p^2(F) = \rho^2$$
- Information approach
  - **Linear relationship** generalizes to **any functional relationship**.
  - **Variance** generalizes to an **uncertainty function**
    (a) Concave function of $f$ that measures the concentration, $H(F) \leq H(\text{uniform})$
    (b) The variance does not always satisfy this condition
  - **The departure from independence** is measured formally by a **divergence function** between two probability distributions $\mathcal{D}(P : Q) \geq 0$, where $\mathcal{D} = 0$ if and only if the distributions are identical $dP(x) = dQ(x)$
  - **Invariance under linear transformations** generalizes to the **invariance under all one-to-one transformations**

# Compare predictability of distributions

- Which of the two distributions ***<u>in each panel</u>*** have outcomes that can be predicted with a high probability?
  - Write your answer for each case as: “solid” or “dashed”
- With which distribution in each panel is more difficult to predict outcomes?



# Compare predictability of distributions



# Association Indices

- Two popular indices
  - Spearman's rank correlation

$$\rho_s(F) = 12 \int\int_{\Re^2} F_1(x_1)F_2(x_2)f(x_1, x_2)dx_1 dx_2 - 3$$

  - Kendall's tau

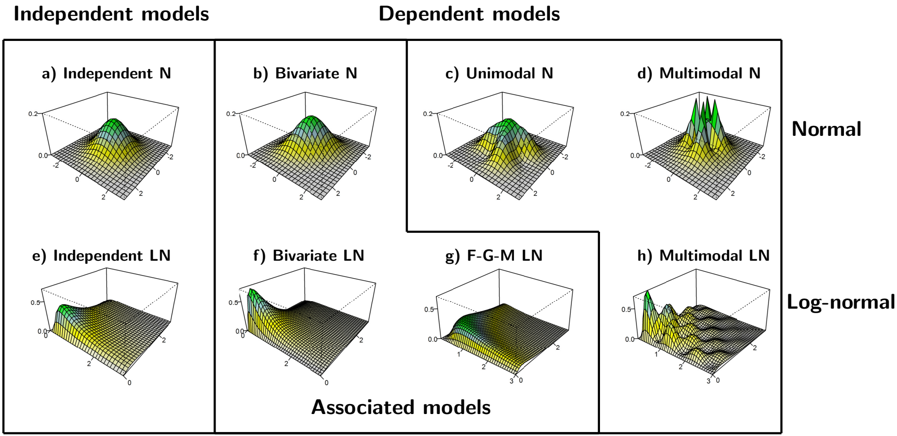
$$\tau(F) = 4 \int\int_{\Re^2} F(x_1, x_2)f(x_1, x_2)dx_1 dx_2 - 1$$

  - Sign indicates the direction of association
  - Invariant under monotone transformations
  - Copula representations

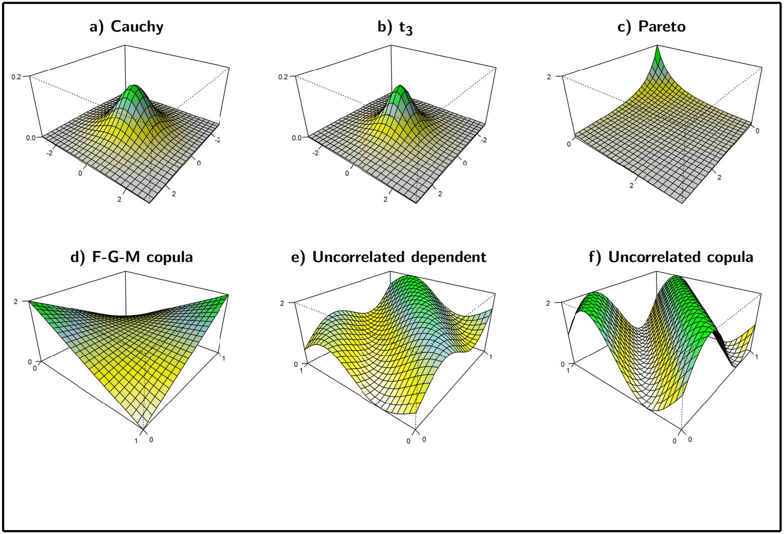
$$\rho_s(F) = \rho_s(C), \quad \tau(F) = \tau(C)$$

  - Two variables are unassociated if and only if $\rho_s = \tau = 0$
  - Unassociated dependent $\rho_s = \tau = 0, \ \delta^2 > 0$
    - One variable is useful for predicting the other
- $\rho_p = \eta^2 = \rho_s = \tau = 0 \ \not\Rightarrow X_1$ and $X_2$ are independent

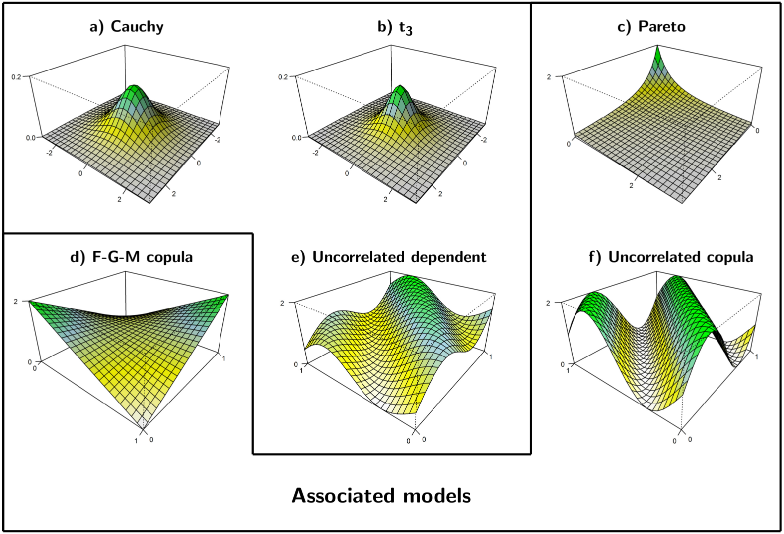
# Example 1. Examples of Bivariate Distributions



# Figure 2. Examples of Bivariate Distributions



# Figure 2. Examples of Bivariate Distributions



# Elliptical & Pareto families

- **Elliptical pdf**

$$f_h(x_1, x_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}}\, h\left(\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right), (x_1, x_2) \in \Re^2, \ \ \rho^2 < 1,$$

$h(\cdot)$ is a real function

  - **Gaussian (Figure 1b)** $h(z) = e^{-z/2}$
  - **Student-t (Figure 2a, $\nu = 3$)** $h(z) = \left(1 + \frac{z}{\nu}\right)^{-\nu/2-1}$
  - **Cauchy (Figure 2b)** $h(z) = (1+z)^{-3/2}$
  - Log-normal (Figure 1f) is monotone transformation of normal

# Elliptical & Pareto families

- **Elliptical pdf**

$$f_h(x_1,x_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}}\, h\left(\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right), (x_1,x_2) \in \Re^2, \;\; \rho^2 < 1,$$

$h(\cdot)$ is a real function

  - **Gaussian (Figure 1b)** $h(z) = e^{-z/2}$
  - **Student-t (Figure 2a, $\boldsymbol{\nu = 3}$)** $h(z) = \left(1 + \frac{z}{\nu}\right)^{-\nu/2-1}$
  - **Cauchy (Figure 2b)** $h(z) = (1+z)^{-3/2}$
  - Log-normal (Figure 1f) is monotone transformation of normal

- **Pareto Type II (Figure 2c)**

$$f(x_1,x_2) = \alpha(\alpha+1)(1+x_1+x_2)^{-\alpha-2}, \;\; x_i \geq 0, \; \alpha > 0$$

Numerous other distributions are related to this model by monotone transformations, including Pareto Types I, III & IV, exponential, Weibull, logistic, Burr, and Calyton copula, among others (Darbellay & Vajda 2000, Asadi et al. 2006, Balakrishnan & Lai 2009)

# Generalized Sarmanov families

- **Families with bivariate pdf's**

$$f_q(x_1, x_2) = f_1(x_1)f_2(x_2)[1 + \beta q(x_1, x_2)], (x_1, x_2) \in \Re^2, \quad \beta \leq B^{-1}$$

  - $f_i(x_i), i = 1, 2$ are the marginal pdf's
  - $q(x_1, x_2)$ is a measurable bounded function $|q(x_1, x_2)| \leq B$

# Generalized Sarmanov families

- **Families with bivariate pdf's**

$$f_q(x_1, x_2) = f_1(x_1)f_2(x_2)[1 + \beta q(x_1, x_2)], (x_1, x_2) \in \Re^2, \quad \beta \leq B^{-1}$$

  - $f_i(x_i), i = 1, 2$ are the marginal pdf's
  - $q(x_1, x_2)$ is a measurable bounded function $|q(x_1, x_2)| \leq B$

- Sarmanov families: $q(x_1, x_2) = q_1(x_1)q_2(x_2)$
  - **F-G-M Copula (Figure 2d, $\beta = 1$)**
    $f_i(x_i) = 1, \ 0 \leq x_i \leq 1, \quad |\beta| \leq 1, \ q_i(x_i) = (1 - 2x_i)$

# Generalized Sarmanov families

- **Families with bivariate pdf's**

$$f_q(x_1,x_2) = f_1(x_1)f_2(x_2)[1+\beta q(x_1,x_2)], (x_1,x_2) \in \Re^2, \quad \beta \le B^{-1}$$

  - $f_i(x_i), i = 1,2$ are the marginal pdf's
  - $q(x_1,x_2)$ is a measurable bounded function $|q(x_1,x_2)| \le B$
- Sarmanov families: $q(x_1,x_2) = q_1(x_1)q_2(x_2)$
  - **F-G-M Copula (Figure 2d, $\boldsymbol{\beta = 1}$)**
    $f_i(x_i) = 1,\ 0 \le x_i \le 1, \quad |\beta| \le 1,\ q_i(x_i) = (1-2x_i)$
  - **Unassociated log-normal (Figure 1h, $\boldsymbol{\beta = 1}$)**
    $f_i(x_i) = LN(0,1),\ x_i > 0, \quad |\beta| \le 1,\ q_i(x_i) = \sin(2\pi \log x_i)$
    - $E(X_i^m|X_j) = E(X_i^m), i \ne j,\ m = 1,2,\cdots$ (De Paula 2008)
    - All polynomial functions of $X_i$ and $X_j$ are uncorrelated

# Generalized Sarmanov families

- A class of models for uncorrelated random variable

$$X_1 + X_2 \stackrel{st}{=} X_2^o + X_2^o, \quad F^o(x_1, x_2) = F_1(x_1)F_2(x_2)$$

Referred as the *Summable Uncorrelated Marginals (SUM)*, (Hamedani & Tata 1975, Ebrahimi et al. 2010)

# Generalized Sarmanov families

- A class of models for uncorrelated random variable

$$X_1 + X_2 \overset{st}{=} X_2^o + X_2^o, \quad F^o(x_1, x_2) = F_1(x_1)F_2(x_2)$$

Referred as the *Summable Uncorrelated Marginals (SUM)*, (Hamedani & Tata 1975, Ebrahimi et al. 2010)

  - **Unassociated normal unimodal (Figure 1c, $\beta = .25e^2$)**
    $f_i(x_i) = N(0,1), \; (x_1, x_2) \in \Re^2, \;\; |\beta| \le .25e^2,$
    $q(x_1, x_2) = x_1 x_2 (x_1^2 - x_2^2) e^{-\frac{1}{2}(x_1^2 + x_2^2)}$
  - **Unassociated normal multimodal (Figure 1d, $\beta = 4$)**
    $f_i(x_i) = N(0,1), \; (x_1, x_2) \in \Re^2, \;\; |\beta| \le 4,$
    $q(x_1, x_2) = x_1 x_2 (x_1^2 - x_2^2) e^{-\frac{1}{2}(x_1^2 + x_2^2)}$
  - **Uncorrelated dependent (Figure 2e, $\beta = .5$)**
    $f_1(x_1) = .5 + x_1, \; f_2(x_2) = 1, \;\; 0 \le x_i \le 1,$
    $|\beta| \le 1, \;\; q(x_1, x_2) = \sin[2\pi(x_2 - x_1)]$
  - **Uncorrelated Copula (Figure 2f, $\beta = .6$)**
    $f_i(x_i) = 1, 0 \le x_i \le 1, \;\; |\beta| \le 1, \; q(x_1, x_2) = \sin[2\pi(x_2 - x_1)]$

# Common metric?

| Family | $\rho_p$ | $\rho_s$ | $\tau$ | $\eta^2$ | $\delta^2$ | Example pdf Figure, $\delta^2$ (Rank*) |
|---|---|---|---|---|---|---|
| **Elliptical** | | | | | | |
| Gaussian | ● | ● | ● | ● | ● | 1b, .160 (7) |
| Student-$t$ | ◓ | ⊙ | ⊙ | ◒ | ● | 2a, .081 (10) |
| Cauchy | ○ | ● | ● | ○ | ● | 2b, .361 (2) |
| **Pareto and related families** | | | | | | |
| Pareto | ◒ | ● | ● | ◒ | ● | 2c, .836 (1) |
| **Generalized Sarmanov** | | | | | | |
| F-G-M copula | ● | ● | ● | ● | ● | 2d, .113 (9) |
| Uncorrelated dependent | ⓪ | ● | ● | ● | ● | 2e, .136 (8) |
| Uncorrelated copula | ⓪ | ⓪ | ● | ● | ● | 2f, .172 (5) |
| Unassociated normal unimodal | ⓪ | ⓪ | ⓪ | ● | ● | 1c, .173 (4) |
| Unassociated normal multimodal | ⓪ | ⓪ | ⓪ | ● | ● | 1d, .168 (6) |
| Unassociated log-normal | ⓪ | ⓪ | ⓪ | ⓪ | ● | 1h, .249 (3) |

Notes:
● suitable within the family; ○ undefined; ⊙ unsuitable, dependence varies within the family, index does not;
⓪ unsuitable, dependence varies, index identically zero; ◒ Partially suitable, undefined for some parameter values;
◓ undefined for some parameter values, dependence varies within the family, index does not;
* ranks are among ten dependent pdf's in Figures 2 and 3.

# Multivariate Information

- $G$ in $K(F : G)$ is a model for independence between two or more subvectors of a $d$-dimensional random vector $\mathbf{X}$
  - The independent model $G = F_1 \cdots F_d$
  - Independence of two disjoint subvectors, $G(\mathbf{x}) = F_j(\mathbf{x}_j)F_h(\mathbf{x}_h), \ j + h = d.$

# Multivariate Information

- $G$ in $K(F:G)$ is a model for independence between two or more subvectors of a $d$-dimensional random vector $\mathbf{X}$
  - The independent model $G = F_1 \cdots F_d$
  - Independence of two disjoint subvectors, $G(\mathbf{x}) = F_j(\mathbf{x}_j)F_h(\mathbf{x}_h),\ j+h=d$.
- Three properties of multivariate $M$'s:

$$M(X_1,\cdots,X_d) \text{ increasing in } d$$

$$M(X_1,(X_2,\cdots,X_d)) = \sum_{i=2}^{d} M(X_1,X_i|X_2,\cdots X_{i-1}) \text{ increasing in } d$$

$$M(\mathbf{X}) = M(\mathbf{X}_j) + M(\mathbf{X}_k) + M(\mathbf{X}_j,\mathbf{X}_h) \geq M(\mathbf{X}_j) + M(\mathbf{X}_h), \quad j+h=d$$

  - $M(X_1,X_i|X_2,\cdots X_{i-1}) = E_{x_2,\cdots x_{i-1}}[M(X_1,X_i|x_2,\cdots x_{i-1})]$, partial mutual information
  - The first two formalize the intuition that dependence increases with the dimension
  - The inequality formalizes the intuition that aggregation of lower dependence underestimates the overall dependence

# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \stackrel{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

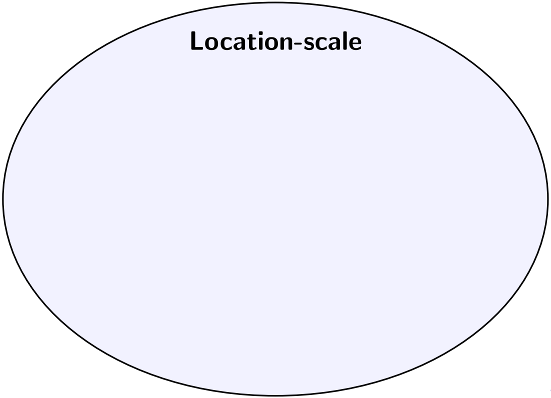
  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.

# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \stackrel{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.

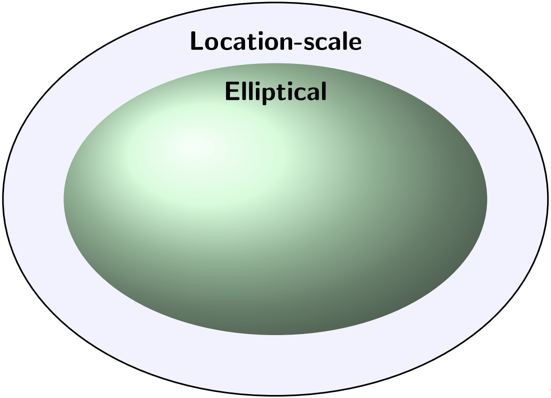


# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \stackrel{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.

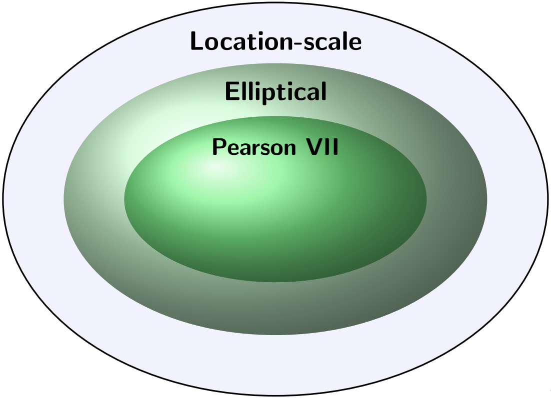


# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \overset{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.

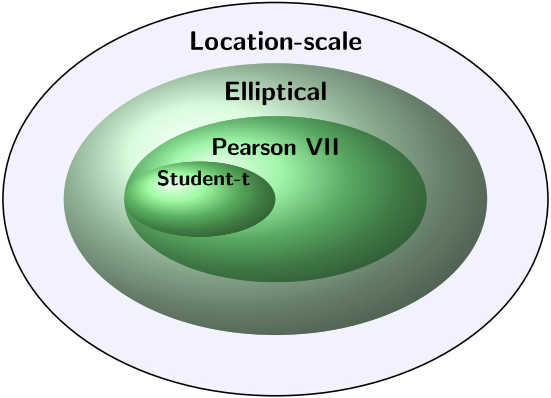


# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \stackrel{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.

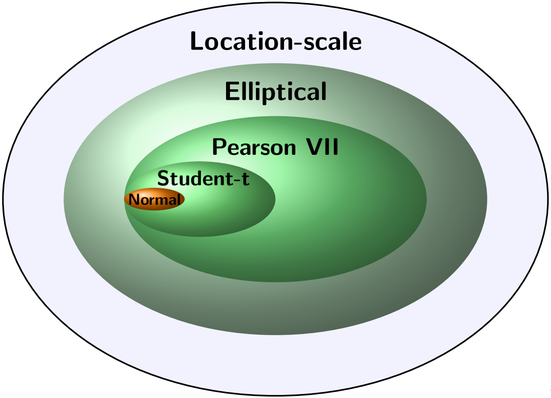


# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \overset{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.

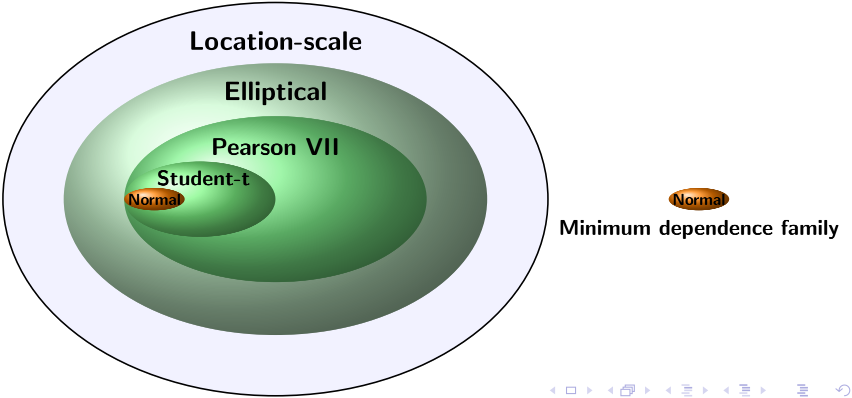


# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \stackrel{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.

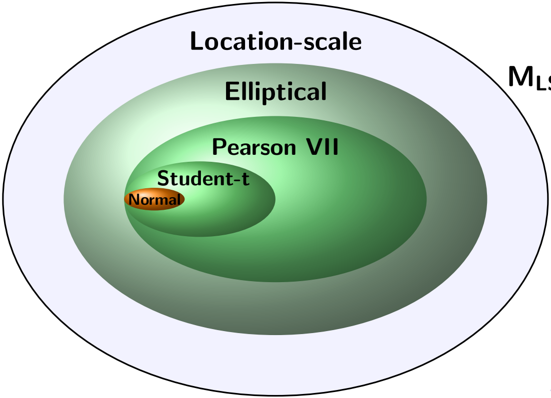


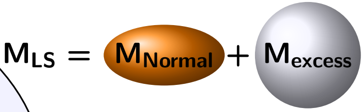
# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \overset{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.



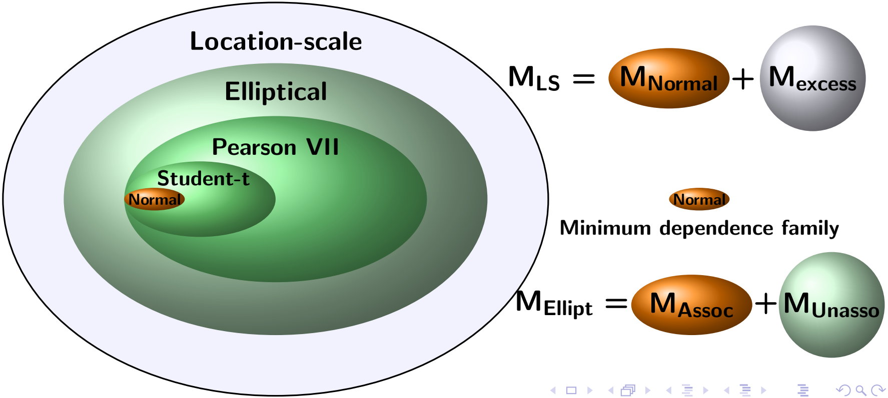




# Location-Scale (L-S) Family

- A pdf $f(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\mu}, \Sigma)$ with location vector $\boldsymbol{\mu}$ and scale matrix $\Sigma$

$$\mathbf{X} \stackrel{st}{=} \boldsymbol{\mu} + |\Sigma|^{1/2}\mathbf{X}^o,$$

  - $\boldsymbol{\theta}$, model parameters other than $\boldsymbol{\mu}$ and $\Sigma$
  - $\mathbf{X}^o$, in the same family with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$, identity matrix.



# Location-Scale Family

- The entropy of L-S family

$$H(\mathbf{X}|\boldsymbol{\theta}, \Sigma) = H(\mathbf{X}^o|\boldsymbol{\theta}) + \frac{1}{2}\log|\Sigma| \leq H_{\mathcal{G}}(\Sigma)$$

  - $H(\mathbf{X}^o|\boldsymbol{\theta})$ is free from $\boldsymbol{\mu}$ and $\Sigma$
  - $H_{\mathcal{G}}(\Sigma)$, the entropy of Gaussian distribution (The maximum entropy model)

# Location-Scale Family

- The entropy of L-S family

$$H(\mathbf{X}|\boldsymbol{\theta}, \Sigma) = H(\mathbf{X}^o|\boldsymbol{\theta}) + \frac{1}{2}\log|\Sigma| \leq H_{\mathcal{G}}(\Sigma)$$

  - $H(\mathbf{X}^o|\boldsymbol{\theta})$ is free from $\boldsymbol{\mu}$ and $\Sigma$
  - $H_{\mathcal{G}}(\Sigma)$, the entropy of Gaussian distribution (The maximum entropy model)
- The mutual information measures

$$M(\mathbf{X}|\boldsymbol{\theta}, \Sigma) = M(\Omega) + M(\mathbf{X}^0|\boldsymbol{\theta}) \geq M(\Omega)$$

  - $\Omega = D^{-1/2}\Sigma D^{-1/2}, \; D = \text{Diag}[\sigma_{11}, \cdots, \sigma_{dd}]$
  - $M(\Omega)$, the portion of dependence induced by the rotation
  - $M(\Omega) = M_{\mathcal{G}}(\Omega)$, for the Gaussian model with correlation $\Omega$
  - $M(\mathbf{X}^0|\boldsymbol{\theta})$, the intrinsic dependence of the unrotated vector $\mathbf{X}^0$

# Location-Scale Family

- The entropy of L-S family

$$H(\mathbf{X}|\boldsymbol{\theta}, \Sigma) = H(\mathbf{X}^o|\boldsymbol{\theta}) + \frac{1}{2}\log|\Sigma| \leq H_{\mathcal{G}}(\Sigma)$$

  - $H(\mathbf{X}^o|\boldsymbol{\theta})$ is free from $\boldsymbol{\mu}$ and $\Sigma$
  - $H_{\mathcal{G}}(\Sigma)$, the entropy of Gaussian distribution (The maximum entropy model)
- The mutual information measures

$$M(\mathbf{X}|\boldsymbol{\theta}, \Sigma) = M(\Omega) + M(\mathbf{X}^0|\boldsymbol{\theta}) \geq M(\Omega)$$

  - $\Omega = D^{-1/2}\Sigma D^{-1/2}, \; D = \text{Diag}[\sigma_{11}, \cdots, \sigma_{dd}]$
  - $M(\Omega)$, the portion of dependence induced by the rotation
  - $M(\Omega) = M_{\mathcal{G}}(\Omega)$, for the Gaussian model with correlation $\Omega$
  - $M(\mathbf{X}^0|\boldsymbol{\theta})$, the intrinsic dependence of the unrotated vector $\mathbf{X}^0$
- Among all distributions in the multivariate L-S family having the same scale matrix $\Sigma$, the Gaussian model (copula) has the minimum dependence model

# Multivariate Normal (Gaussian) Information Measures

| Mutual Information | $M_{\mathcal{G}}(\Omega)$ | Index $(\delta^2_{\mathcal{G}})$ |
|---|---|---|
| $M(\mathbf{X})$ | $-.5\log\lvert\Omega\rvert$ | $1-\lvert\Omega\rvert$ |
| $M(\mathbf{X}_1,\mathbf{X}_2)$ | $-.5\sum_{j=1}^{\min\{d_k\}}\log(1-\lambda_j)$ | $1-\prod_{j=1}^{\min\{d_k\}}(1-\lambda_j)$ |
| $M[Y,(X_1,\cdots,X_d)]$ | $-.5\log\left(1-\rho^2_{y\lvert x_1,\cdots,x_d}\right)$ | $\rho^2_{y\lvert x_1,\cdots,x_d}$ |

- Row 1. measures for shared information between all components.
- Row 2. measures for two disjoint subvectors
  - $\lambda_j,\ j=1,\cdots,\min\{d_k\}$ are the nonzero eigenvalues of $\Omega_{11}^{-1/2}\Omega_{12}\Omega_{22}^{-1/2}\Omega_{12}\Omega_{11}^{-1/2}$ [$\Omega_{ij}$ partitions of $\Omega$ for $(\mathbf{X}_1,\mathbf{X}_2)$]
  - The canonical correlations of the two subvectors $(\mathbf{X}_1,\mathbf{X}_2)$
- Row 3. regression information measures
  - $\rho^2_{y\lvert x_1,\cdots,x_d}$, the normal regression fit index

# Elliptical families

- The pdf is in the form of

$$f_h(\mathbf{x}|\Sigma, \boldsymbol{\mu}) = k|\Sigma|^{-1/2} h\Big((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big)$$

  - $h(\cdot)$ is referred to as the scale or generator function which may include a vector of parameters $\boldsymbol{\theta}$ in addition to $(\boldsymbol{\mu}, \Sigma)$
  - The marginal distributions are also elliptical with L-S parameters $(\mu_i, \sigma_{ii})$, but the generator of the marginals $h_i(\cdot)$ may be different than $h(\cdot)$.

# Elliptical families

- The pdf is in the form of

$$f_h(\mathbf{x}|\Sigma, \boldsymbol{\mu}) = k|\Sigma|^{-1/2} h\Big((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big)$$

  - $h(\cdot)$ is referred to as the scale or generator function which may include a vector of parameters $\boldsymbol{\theta}$ in addition to $(\boldsymbol{\mu}, \Sigma)$
  - The marginal distributions are also elliptical with L-S parameters $(\mu_i, \sigma_{ii})$, but the generator of the marginals $h_i(\cdot)$ may be different than $h(\cdot)$.
- The Gaussian family $N(\boldsymbol{\mu}, \Sigma)$: $h(z) = e^{z^2/2}$

# Elliptical families

- The pdf is in the form of

$$f_h(\mathbf{x}|\Sigma, \boldsymbol{\mu}) = k|\Sigma|^{-1/2} h\Big((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big)$$

  - $h(\cdot)$ is referred to as the scale or generator function which may include a vector of parameters $\boldsymbol{\theta}$ in addition to $(\boldsymbol{\mu}, \Sigma)$
  - The marginal distributions are also elliptical with L-S parameters $(\mu_i, \sigma_{ii})$, but the generator of the marginals $h_i(\cdot)$ may be different than $h(\cdot)$.
- The Gaussian family $N(\boldsymbol{\mu}, \Sigma)$: $h(z) = e^{z^2/2}$
- Student-$t$ family $t(\nu, \boldsymbol{\mu}, \Sigma)$, $\nu$, degrees of freedom:

$$h(z) = \left(1 + \frac{z^2}{\nu}\right)^{-(d+\nu)/2}$$

# Elliptical families

- The pdf is in the form of

$$f_h(\mathbf{x}|\Sigma, \boldsymbol{\mu}) = k|\Sigma|^{-1/2} h\Big((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big)$$

  - $h(\cdot)$ is referred to as the scale or generator function which may include a vector of parameters $\boldsymbol{\theta}$ in addition to $(\boldsymbol{\mu}, \Sigma)$
  - The marginal distributions are also elliptical with L-S parameters $(\mu_i, \sigma_{ii})$, but the generator of the marginals $h_i(\cdot)$ may be different than $h(\cdot)$.
- The Gaussian family $N(\boldsymbol{\mu}, \Sigma)$: $h(z) = e^{z^2/2}$
- Student-$t$ family $t(\nu, \boldsymbol{\mu}, \Sigma)$, $\nu$, degrees of freedom:

$$h(z) = \left(1 + \frac{z^2}{\nu}\right)^{-(d+\nu)/2}$$

- For all elliptical families: $\tau = \dfrac{2}{\pi}\sin^{-1}(\rho)$; (Fang et al. 2002)

# Multivariate Student-$t$

- Relationships with multivariate normal
  - Limiting distribution: $t(\nu, \boldsymbol{\mu}, \Sigma) \to N(\boldsymbol{\mu}, \Sigma)$ as $\nu \to \infty$

# Multivariate Student-$t$

- Relationships with multivariate normal
  - Limiting distribution: $t(\nu, \boldsymbol{\mu}, \Sigma) \to N(\boldsymbol{\mu}, \Sigma)$ as $\nu \to \infty$
  - Normal mixture:

$$f(\mathbf{x}|\phi) = N(\boldsymbol{\mu}, \phi\Sigma), \quad \phi \sim \text{Gamma}(\nu/2, \nu/2),\ \nu = 1, 2 \cdots$$
$$\Longrightarrow \quad f(\mathbf{x}) = t(\nu, \boldsymbol{\mu}, \Sigma)$$

# Multivariate Student-$t$

- Relationships with multivariate normal
  - Limiting distribution: $t(\nu, \boldsymbol{\mu}, \Sigma) \to N(\boldsymbol{\mu}, \Sigma)$ as $\nu \to \infty$
  - Normal mixture:

$$f(\mathbf{x}|\phi) = N(\boldsymbol{\mu}, \phi\Sigma), \quad \phi \sim \text{Gamma}(\nu/2, \nu/2), \ \nu = 1, 2 \cdots$$
$$\Longrightarrow \quad f(\mathbf{x}) = t(\nu, \boldsymbol{\mu}, \Sigma)$$

  - The information index:

$$\delta_T(\nu, \Sigma) = \delta_{\mathcal{G}} + (1 - \delta_{\mathcal{G}})\delta(\nu), \quad \uparrow \boldsymbol{\delta_{\mathcal{G}}}, \quad \downarrow \boldsymbol{\nu}$$

    - $\delta_{\mathcal{G}}$ the index for Gaussian (normal)
    - For the bivariate case:

$$\delta_t^2(\nu, \tau) = \sin^2\left(\frac{\tau\pi}{2}\right) + \cos^2\left(\frac{\tau\pi}{2}\right)\delta^2(\nu),$$

# Multivariate Student-$t$

- Relationships with multivariate normal
  - Limiting distribution: $t(\nu, \boldsymbol{\mu}, \Sigma) \to N(\boldsymbol{\mu}, \Sigma)$ as $\nu \to \infty$
  - Normal mixture:

$$f(\mathbf{x}|\phi) = N(\boldsymbol{\mu}, \phi\Sigma), \quad \phi \sim \text{Gamma}(\nu/2, \nu/2), \ \nu = 1, 2 \cdots$$
$$\Longrightarrow \quad f(\mathbf{x}) = t(\nu, \boldsymbol{\mu}, \Sigma)$$

  - The information index:

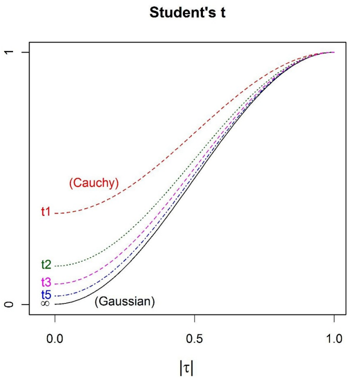
$$\delta_T(\nu, \Sigma) = \delta_{\mathcal{G}} + (1 - \delta_{\mathcal{G}})\delta(\nu), \quad \uparrow \boldsymbol{\delta_{\mathcal{G}}}, \quad \downarrow \boldsymbol{\nu}$$

    - $\delta_{\mathcal{G}}$ the index for Gaussian (normal)
    - For the bivariate case:

$$\delta_t^2(\nu, \tau) = \sin^2\left(\frac{\tau\pi}{2}\right) + \cos^2\left(\frac{\tau\pi}{2}\right)\delta^2(\nu),$$

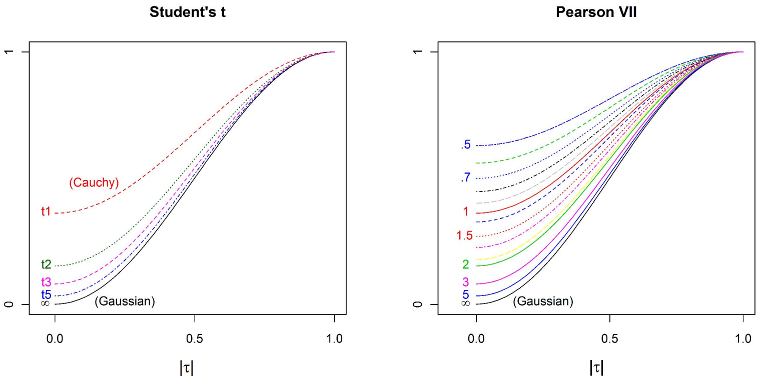
- Many applications
  - Regression model with multivariate $t$ errors (Zellner 1976)
  - Dynamic stochastic general equilibrium model with Student-$t$ errors (Chib & Ramamurthy 2012)
  - Copulas (Demarta & McNeil 2005), model for financial variables, ...

# Student-$t$




- Student-$t$ when $\tau$ ($\rho$) and the degrees of freedom $\nu$ are low
  - Substantial gaps between the level of dependence
  - The spectrum of dependence is narrow

# Student's $t$ & Pearson Type VII




- Student-$t$ when $\tau$ ($\rho$) and the degrees of freedom $\nu$ are low
  - Substantial gaps between the level of dependence
  - The spectrum of dependence is narrow
- The spectrum of dependence of the $t$ family is substantially widen and refined by replacing $\nu/2$ with a parameter $\alpha > 0$.

# Convolution Models

- Noisy relationship between $Y$ and $\mathbf{X}' = (X_1, \cdots, X_p)$:

$$Y = \phi(\mathbf{X}, \boldsymbol{\beta}) + \epsilon$$

  - $\phi(\cdot, \cdot)$, a scalar function, need not be linear
  - $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$,
  - $\mathbf{X}$, $\boldsymbol{\beta}$, or both can be stochastic
  - $\epsilon$, random noise, may not be independent of the signal $\phi(\mathbf{X}_i, \boldsymbol{\beta})$

# Convolution Models

- Noisy relationship between $Y$ and $\mathbf{X}' = (X_1, \cdots, X_p)$:

$$Y = \phi(\mathbf{X}, \boldsymbol{\beta}) + \epsilon$$

  - $\phi(\cdot, \cdot)$, a scalar function, need not be linear
  - $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$,
  - $\mathbf{X}$, $\boldsymbol{\beta}$, or both can be stochastic
  - $\epsilon$, random noise, may not be independent of the signal $\phi(\mathbf{X}_i, \boldsymbol{\beta})$
- Two known results for the sum of independent variables
  - The sum has a larger entropy than each variable alone
  - The amount of increase in the entropy is the $M$ between the sum and the summands (Blahut 1987)

# Convolution Models

- Noisy relationship between $Y$ and $\mathbf{X}' = (X_1, \cdots, X_p)$:

$$Y = \phi(\mathbf{X}, \boldsymbol{\beta}) + \epsilon$$

  - $\phi(\cdot, \cdot)$, a scalar function, need not be linear
  - $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$,
  - $\mathbf{X}$, $\boldsymbol{\beta}$, or both can be stochastic
  - $\epsilon$, random noise, may not be independent of the signal $\phi(\mathbf{X}_i, \boldsymbol{\beta})$
- Two known results for the sum of independent variables
  - The sum has a larger entropy than each variable alone
  - The amount of increase in the entropy is the $M$ between the sum and the summands (Blahut 1987)
- An enhanced version of Blahut's result gives:

$$M(Y, \phi(\mathbf{X}, \boldsymbol{\beta})) = [H(Y) - H(\epsilon)] + M(\epsilon, \phi(\mathbf{X}, \boldsymbol{\beta}))$$

- Blahut's theorem is for $M(\epsilon, \phi(\mathbf{X}, \boldsymbol{\beta})) = 0$:

$$M(Y, \phi(\mathbf{X}, \boldsymbol{\beta})) = H(Y) - H(\epsilon)$$

# Normal Linear Regression

- $\phi(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}'\boldsymbol{\beta}, \; Y|\mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}'\boldsymbol{\beta}, \sigma^2), \; M(\epsilon, \phi(\mathbf{X}, \boldsymbol{\beta})) = 0$
- Two information measures are well-known

# Normal Linear Regression

- $\phi(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}'\boldsymbol{\beta}, \;\; Y|\mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}'\boldsymbol{\beta}, \sigma^2), \; M(\epsilon, \phi(\mathbf{X}, \boldsymbol{\beta})) = 0$
- Two information measures are well-known
  - $\boldsymbol{\beta}$ non-stochastic:

$$\begin{aligned} M(Y, \phi(\mathbf{X}, \boldsymbol{\beta})) = M(Y, \mathbf{X}) &= -\frac{1}{2}\log(1 - \rho^2_{y|x_1,\cdots x_p}) \\ &= -\frac{1}{2}\sum_{j=1}^{p}\log(1 - \rho^2_{x_j|x_1,\cdots x_{j-1}}), \end{aligned}$$

    - $\rho^2_{x_j|x_1,\cdots x_{j-1}}$ is the squared partial correlation coefficient
    - Theil and Chung (1988) proposed the above decomposition of the sample version $2\hat{M}(\mathbf{Y}, \mathbf{X}) = -\log(1 - R^2)$ as transformation of the regression index $R^2$ for assessing the relative importance of the predictors.

# Normal Linear Regression

- $\phi(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}'\boldsymbol{\beta}, \;\; Y|\mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}'\boldsymbol{\beta}, \sigma^2), \; M(\epsilon, \phi(\mathbf{X}, \boldsymbol{\beta})) = 0$
- Two information measures are well-known
  - $\boldsymbol{\beta}$ non-stochastic:

$$\begin{aligned} M(Y, \phi(\mathbf{X}, \boldsymbol{\beta})) = M(Y, \mathbf{X}) &= -\frac{1}{2}\log(1 - \rho^2_{y|x_1, \cdots x_p}) \\ &= -\frac{1}{2}\sum_{j=1}^{p}\log(1 - \rho^2_{x_j|x_1, \cdots x_{j-1}}), \end{aligned}$$

    - $\rho^2_{x_j|x_1, \cdots x_{j-1}}$ is the squared partial correlation coefficient
    - Theil and Chung (1988) proposed the above decomposition of the sample version $2\hat{M}(\mathbf{Y}, \mathbf{X}) = -\log(1 - R^2)$ as transformation of the regression index $R^2$ for assessing the relative importance of the predictors.
  - $\mathbf{X}$ a design matrix (non-stochastic) and $\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\mu}_0, \sigma_0^2 A_0)$:

$$M(\mathbf{Y}, \boldsymbol{\beta}|\eta, A_0) = \frac{1}{2}\log\left|I_p + \frac{\sigma_0^2}{\sigma^2}A_0 X'X\right|.$$

  The Bayesian sample information about regression parameter (Lindley's information)

# Bayesian Linear Regression Beyond Normal

- $\epsilon \sim F(0, \sigma^2)$
- $\beta_j, j = 1, \cdots, p$ are independent and have $g$-priors $\beta_j \sim F(0, \sigma^2 x_j^{-1}),\ j = 1, \cdots, p$
- $Y_i$ is convolution of $p + 1$ iid variables $Z_j = \beta_j x_j$ and $\epsilon$
- If $F$ is closed under convolution, $Y_i | \sigma^2 \sim F(0, (p+1)\sigma^2)$

# Bayesian Linear Regression Beyond Normal

- $\epsilon \sim F(0, \sigma^2)$
- $\beta_j, j = 1, \cdots, p$ are independent and have $g$-priors
  $\beta_j \sim F(0, \sigma^2 x_j^{-1}),\ j = 1, \cdots, p$
- $Y_i$ is convolution of $p+1$ iid variables $Z_j = \beta_j x_j$ and $\epsilon$
- If $F$ is closed under convolution, $Y_i|\sigma^2 \sim F(0, (p+1)\sigma^2)$
- Information quantities when $F$ are normal and Cauchy
  - Regression with $t$ error has been proposed for capturing outliers (Zellner 1976, Lang et al. 1989)

| $F$ | $H(Y)$ | $H(\epsilon)$ | $M(Y, \boldsymbol{\beta})$ | $\delta^2(Y, \boldsymbol{\beta})$ |
|---|---|---|---|---|
| Normal | $.5 \log(2(p+1)\pi e\sigma^2)$ | $.5 \log(2\pi e\sigma^2)$ | $.5 \log(p+1)$ | $\frac{p}{p+1}$ |
| Cauchy | $\log(4(p+1)\pi\sigma^2)$ | $\log(4\pi\sigma^2)$ | $\log(p+1)$ | $\frac{p}{p+1} + \frac{p}{(p+1)^2}$ |

# Stochastic Process

- $T_1, T_2, \cdots$ inter-arrival times of failures of a repairable system
- Time to the $n$th failure, $Y_n = \sum_{i=1}^{n} T_i$
  - Distribution of the $i$th failure time is gamma, $\text{Ga}(\alpha_i, \lambda)$
  - Failures are independent
  - Distribution of $Y_n$ is $\text{Ga}(\beta_n, \lambda)$, $\beta_n = \sum_{j=1}^{n} \alpha_i$
- The entropy of $\text{Ga}(\beta_k, 1)$

$$H_G(Y_k) = \log \Gamma(\beta_k) - (\beta_k - 1)\psi(\beta_k) + \beta_k$$

- The convolution result for the independent case is applicable

$$M(Y_n, Y_{n+k}) = H_G(\beta_{n+k}) - H_G(\beta_k),$$

  - $M(Y_n, Y_{n+k})$ is the mutual information of the McKay's bivariate gamma distribution with parameters $(\beta_k, \beta_{n-k}, \lambda)$
  - For the important case of Poisson process, $\beta_k = k$.

# Convolution Models: Dependent Components

- Applications include:
  - Endogenous regression
  - Measurement error models

$$Y = X + \epsilon, \quad X \sim N(\mu, \sigma^2), \;\; \epsilon \sim N(0, \theta^2)$$

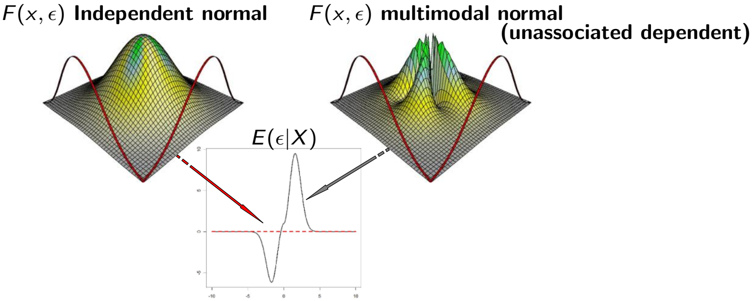
  - $E(X\epsilon) = 0$ ($X$ and $\epsilon$ uncorrelated)

# Convolution Models: Dependent Components

- Applications include:
  - Endogenous regression
  - Measurement error models

$$Y = X + \epsilon, \quad X \sim N(\mu, \sigma^2), \;\; \epsilon \sim N(0, \theta^2)$$

  - $E(X\epsilon) = 0$ ($X$ and $\epsilon$ uncorrelated)



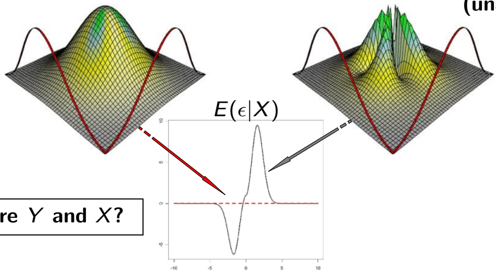
# Convolution Models: Dependent Components

- Applications include:
  - Endogenous regression
  - Measurement error models

$$Y = X + \epsilon, \quad X \sim N(\mu, \sigma^2), \quad \epsilon \sim N(0, \theta^2)$$

  - $E(X\epsilon) = 0$ ($X$ and $\epsilon$ uncorrelated)

$F(x, \epsilon)$ **Independent normal** $F(x, \epsilon)$ **multimodal normal (unassociated dependent)**




**How dependent are $Y$ and $X$?**

# Convolution Models: Dependent Components

- The easy case: $F(x, \epsilon)$ bivariate normal with correlation $\rho$
  - The joint distribution of $(Y, X)$ is also bivariate normal with correlation $\sqrt{.5(1+\rho)}$

$$M(Y, X) = .5 \log 2 - .5 \log(1-\rho)$$

  - More (less) than the independent case when $\rho > 0$ ($\rho < 0$)

# Convolution Models: Dependent Components

- The easy case: $F(x, \epsilon)$ bivariate normal with correlation $\rho$
  - The joint distribution of $(Y, X)$ is also bivariate normal with correlation $\sqrt{.5(1+\rho)}$

$$M(Y, X) = .5 \log 2 - .5 \log(1 - \rho)$$

  - More (less) than the independent case when $\rho > 0$ ($\rho < 0$)
- It is also easy to calculate $H(Y)$ and $H(\epsilon)$ when $F(x, \epsilon)$ is Cauchy or F-G-M copula
- In general, direct computation is tedious

# A Class of Models for Uncorrelated Variables

- **Summable Uncorrelated Marginals** (Ebrahimi et al. 2010c)
  - Defined by the **stochastic equality** $Z_1 + Z_2 \stackrel{st}{=} Z_1^* + Z_2^*$

$$F^*(z_1, z_2) = F_1(z_1)F_2(z_2) \Longrightarrow H(Z_1 + Z_2) = H(Z_1^* + Z_2^*)$$

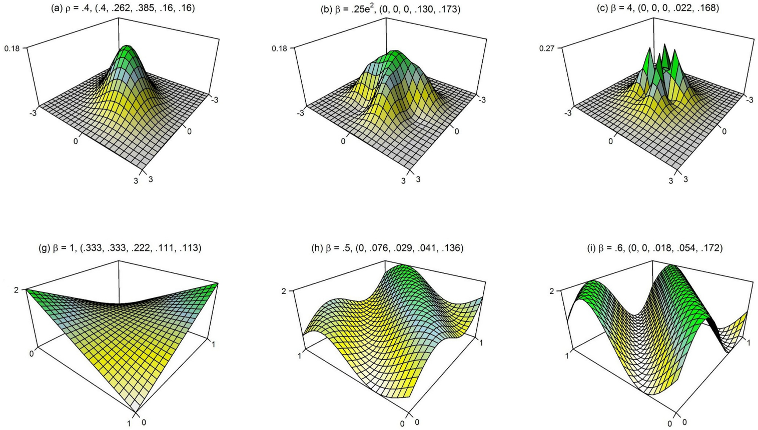
  - ANOVA type decomposition of dependence

$$M(Y, Z_i) = M(Y, Z_i^*) + M(Z_1, Z_2)$$

    - $Y = Z_1 + Z_2$
    - $M(Y, Z_i^*) = H(Y) - H(Z_j), i \neq j = 1, 2$
      (the independent case)

# Examples

$\delta^2(\mathsf{Y}, \mathsf{X}) = .700$ $\delta^2(\mathsf{Y}, \mathsf{X}) = .588$ $\delta^2(\mathsf{Y}, \mathsf{X}) = .591$



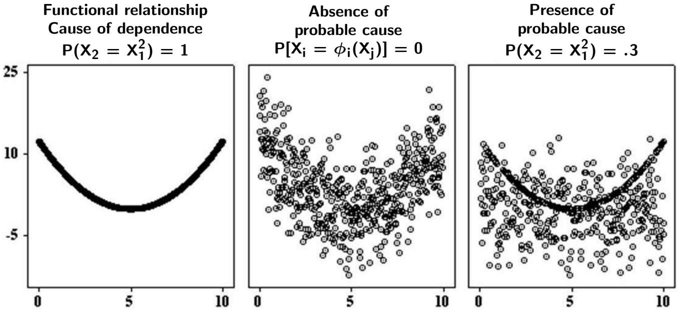

$\delta^2(\mathsf{Y}, \mathsf{X}) = .733$ $\delta^2(\mathsf{Y}, \mathsf{X}) = .616$ $\delta^2(\mathsf{Y}, \mathsf{X}) = .677$

# Absence & Presence of a Probable Cause

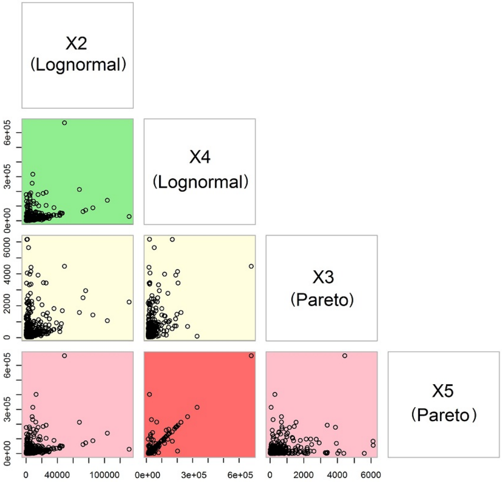
- **“Cause of dependence”**, a functional relationship enabling the perfect predictability
- **"Probable cause”**, a legal terminology for a condition that calls for prudence
  - Absence of a probable cause $P[X_i = \phi_i(X_j)] = 0$,
    **(absolutely continuous distributions)**
  - Presence of a probable cause $P[X_i = \phi_i(X_j)] > 0$,
    **(singular distributions)**

# Absence & Presence of a Probable Cause

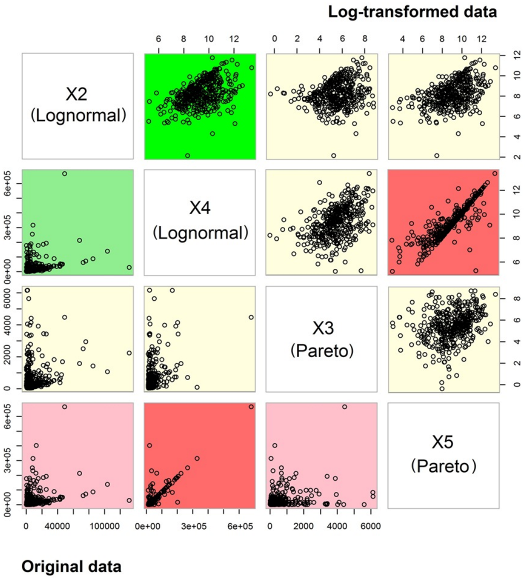
- **“Cause of dependence”**, a functional relationship enabling the perfect predictability
- **”Probable cause”**, a legal terminology for a condition that calls for prudence
  - Absence of a probable cause $P[X_i = \phi_i(X_j)] = 0$, **(absolutely continuous distributions)**
  - Presence of a probable cause $P[X_i = \phi_i(X_j)] > 0$, **(singular distributions)**



# Scatter Plots of Data from a Financial Institution



# Scatter Plots of Data from a Financial Institution



# Singular distributions: Probable cause of dependence

- One variable is not completely dependent on the other
- A functional dependence is probable,

$$0 < P[X_i = \phi_i(X_j)] = \pi < 1$$

  - **Probable cause of dependence**
- The joint distribution $F(x_1, x_2)$ is singular.
- The survival function has the following representation:

$$\bar{F}(x_1, x_2) = (1 - \pi)\bar{F}_a(x_1, x_2) + \pi \bar{F}_s(x_1, x_2),$$

  - $\bar{F}_a(x_1, x_2)$ is the survival function with an absolutely continuous bivariate pdf $f_a(x_1, x_2)$,
  - $\bar{F}_s(x_1, x_2)$ is the survival for a singular part with a univariate pdf $f_s(x)$, $x_i = \phi_i(x_j)$

$$\pi = \int f_s(x)dx$$

# Singular distributions: Some Applications

- Shock models (Marshall & Olkin, 1967)
  - A system with two components $C_i$
    - Three types of shocks $S_j, 1 = 1, 2, 3$
    - $S_j$ kills $C_j, j = 1, 2$ and shock $S_3$ kills both components
  - Marshall-Olkin Bivariate Exponential (MOBE) distribution

# Singular distributions: Some Applications

- Shock models (Marshall & Olkin, 1967)
  - A system with two components $C_i$
    - Three types of shocks $S_j, 1 = 1, 2, 3$
    - $S_j$ kills $C_j, j = 1, 2$ and shock $S_3$ kills both components
  - Marshall-Olkin Bivariate Exponential (MOBE) distribution
- Bayesian test of sharp hypothesis (e.g., Bernardo & Rueda 2002)

# Singular distributions: Some Applications

- Shock models (Marshall & Olkin, 1967)
  - A system with two components $C_i$
    - Three types of shocks $S_j, 1 = 1, 2, 3$
    - $S_j$ kills $C_j, j = 1, 2$ and shock $S_3$ kills both components
  - Marshall-Olkin Bivariate Exponential (MOBE) distribution
- Bayesian test of sharp hypothesis (e.g., Bernardo & Rueda 2002)
- The first order exponential autoregressive (Gaver & Lewis, 1980)

$$X_{n+1} = \rho X_n + \epsilon_{n+1},$$

  - $\{X_n\}$ is a sequence of identically distributed exponential random variables $P(X_n > x) = \bar{F}(x) = e^{-\lambda x}$
  - $\{\epsilon_n\}$ is an iid sequence, $\epsilon_{n+1}$ and $X_n$ are independent
  - $F(x_n, x_{n+1})$ is singular $\pi = \rho$
  - Pareto process (Yeh, et al. 1988), transformation of $X_n$

# Singular distributions: Some Applications

- Shock models (Marshall & Olkin, 1967)
  - A system with two components $C_i$
    - Three types of shocks $S_j, 1 = 1, 2, 3$
    - $S_j$ kills $C_j, j = 1, 2$ and shock $S_3$ kills both components
  - Marshall-Olkin Bivariate Exponential (MOBE) distribution
- Bayesian test of sharp hypothesis (e.g., Bernardo & Rueda 2002)
- The first order exponential autoregressive (Gaver & Lewis, 1980)

$$X_{n+1} = \rho X_n + \epsilon_{n+1},$$

  - $\{X_n\}$ is a sequence of identically distributed exponential random variables $P(X_n > x) = \bar{F}(x) = e^{-\lambda x}$
  - $\{\epsilon_n\}$ is an iid sequence, $\epsilon_{n+1}$ and $X_n$ are independent
  - $F(x_n, x_{n+1})$ is singular $\pi = \rho$
  - Pareto process (Yeh, et al. 1988), transformation of $X_n$
- Importance of s component for a system: Bivariate distribution of a component's lifetime $X_i,\ i = 1, \cdots, n$ and system's lifetime given by any one of the order statistics $Y_1 \leq \cdots, \leq Y_n$ (Ebrahimi, Jalali, Soofi, & Soyer, forthcoming)

# Information Index for Singular distributions

- $M$ is not defined since $F \not\ll F_1 F_2$

# Information Index for Singular distributions

- $M$ is not defined since $F \not\ll F_1 F_2$
- Generalized information index

$$\delta_\pi(X_1, X_2) = \pi + (1 - \pi)\delta_a(X_1, X_2)$$

  - $\delta_a(X_1, X_2)$ is the dependence information index for the absolutely continuous distribution $\bar{F}_a$

# Information Index for Singular distributions

- $M$ is not defined since $F \not\ll F_1 F_2$
- Generalized information index

$$\delta_\pi(X_1, X_2) = \pi + (1-\pi)\delta_a(X_1, X_2)$$

  - $\delta_a(X_1, X_2)$ is the dependence information index for the absolutely continuous distribution $\bar{F}_a$
  - $\delta_a(X_1, X_2) = 0$ if and only if the variables are independent
  - $\delta_\pi(X_1, X_2) = 1$ if $\pi = 1$
  - $\delta_\pi(X_1, X_2) = \delta_a(X_1, X_2)$ if $\pi = 0$, the case when $F \ll F_1 F_2$

# Information Index for Singular distributions

- $M$ is not defined since $F \not\ll F_1 F_2$
- Generalized information index

$$\delta_\pi(X_1, X_2) = \pi + (1 - \pi)\delta_a(X_1, X_2)$$

  - $\delta_a(X_1, X_2)$ is the dependence information index for the absolutely continuous distribution $\bar{F}_a$
  - $\delta_a(X_1, X_2) = 0$ if and only if the variables are independent
  - $\delta_\pi(X_1, X_2) = 1$ if $\pi = 1$
  - $\delta_\pi(X_1, X_2) = \delta_a(X_1, X_2)$ if $\pi = 0$, the case when $F \ll F_1 F_2$
- Invariant under one-to-one transformations of each variable

# Information Index for Singular distributions

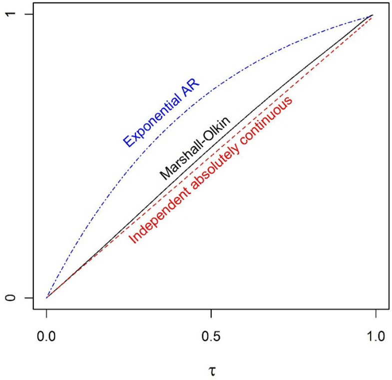
- $M$ is not defined since $F \not\ll F_1 F_2$
- Generalized information index

$$\delta_\pi(X_1, X_2) = \pi + (1-\pi)\delta_a(X_1, X_2)$$

  - $\delta_a(X_1, X_2)$ is the dependence information index for the absolutely continuous distribution $\bar{F}_a$
  - $\delta_a(X_1, X_2) = 0$ if and only if the variables are independent
  - $\delta_\pi(X_1, X_2) = 1$ if $\pi = 1$
  - $\delta_\pi(X_1, X_2) = \delta_a(X_1, X_2)$ if $\pi = 0$, the case when $F \ll F_1 F_2$
- Invariant under one-to-one transformations of each variable
- Based on:
  - the partition property of information
  - applying the probabilistic argument of Marshall and Olkin (1967) to dependence between $X_1$ and $X_2$

$$P[X_i = \phi_i(X_j)] = \pi > 0$$
$$P[X_i \neq \phi_i(X_j)] = 1 - \pi > 0$$

# Examples




- Independent exponential with a singular part included ($\pi = \rho = \tau$)
- The Marshall-Olkin Bivariate Exponential ($\pi = \rho = \tau$)
- The exponential autoregressive ($\pi = \rho = \tau$)

# Bayesian Test of Sharp Hypothesis

- Two parameters $(\theta_1, \theta_2) \in \mathcal{S}$, a continuous region in $\Re^2$
- Test $H_1 : \theta_i = \alpha\theta_j,\ j \neq i = 1, 2$ against $H_1 : \theta_i \neq \alpha\theta_j,\ j \neq i = 1, 2$.
- The plausibility of $H_1$ is described by prior probability $P(H_1) = \pi$.

# Bayesian Test of Sharp Hypothesis

- Two parameters $(\theta_1, \theta_2) \in \mathcal{S}$, a continuous region in $\Re^2$
- Test $H_1 : \theta_i = \alpha\theta_j,\ j \neq i = 1, 2$ against $H_1 : \theta_i \neq \alpha\theta_j,\ j \neq i = 1, 2$.
- The plausibility of $H_1$ is described by prior probability $P(H_1) = \pi$.
- The joint prior distribution $P(\theta_1, \theta_2),\ (\theta_1, \theta_2) \in \mathcal{S}$ has a singular part with a univariate pdf $p_s(\theta_i)$ for $\theta_i = \alpha\theta_j$ and an absolutely continuous part with a bivariate pdf $p_a(\theta_1, \theta_2),\ (\theta_1, \theta_2) \in \mathcal{S}$.

# Bayesian Test of Sharp Hypothesis

- Two parameters $(\theta_1, \theta_2) \in \mathcal{S}$, a continuous region in $\Re^2$
- Test $H_1 : \theta_i = \alpha\theta_j,\ j \neq i = 1, 2$ against $H_1 : \theta_i \neq \alpha\theta_j,\ j \neq i = 1, 2$.
- The plausibility of $H_1$ is described by prior probability $P(H_1) = \pi$.
- The joint prior distribution $P(\theta_1, \theta_2),\ (\theta_1, \theta_2) \in \mathcal{S}$ has a singular part with a univariate pdf $p_s(\theta_i)$ for $\theta_i = \alpha\theta_j$ and an absolutely continuous part with a bivariate pdf $p_a(\theta_1, \theta_2),\ (\theta_1, \theta_2) \in \mathcal{S}$.
- The posterior distribution $P(\theta_1, \theta_2 | D)$ is also singular
- The updated probability of the singularity is given by

$$\pi^* = P(H_1|D) = \frac{\pi}{\pi + (1-\pi)B_{21}}.$$

$B_{21}$, the Bayes factor

# Example: Test equality of two normal means

- $f(\mathbf{x}|\theta_1, \theta_2, \phi\Omega) = N\Big((\theta_1, \theta_2), \phi\Omega\Big)$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$

# Example: Test equality of two normal means

- $f(\mathbf{x}|\theta_1, \theta_2, \phi\Omega) = N\Big((\theta_1, \theta_2), \phi\Omega\Big)$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$
- Normal-gamma prior under $H_2$

$$P(\theta_1, \theta_2|\phi, H_2) = N\Big((m_1, m_2), h\phi\Omega\Big)$$
$$P(\phi) = \text{Ga}(\nu/2, \nu/2)$$

# Example: Test equality of two normal means

- $f(\mathbf{x}|\theta_1, \theta_2, \phi\Omega) = N\Big((\theta_1, \theta_2), \phi\Omega\Big)$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$
- Normal-gamma prior under $H_2$

$$P(\theta_1, \theta_2|\phi, H_2) = N\Big((m_1, m_2), h\phi\Omega\Big)$$
$$P(\phi) = \text{Ga}(\nu/2, \nu/2)$$

| | $P(\theta_1, \theta_2\|\phi)$ | $P(\theta_1, \theta_2)$ |
|---|---|---|
| Absolutely continuous part | Bivariate normal$(\rho)$ | Bivariate $t(\rho, \nu)$ |
| $\delta^2_a(\Theta_1, \Theta_2)$ | $\rho^2$ | $\delta^2_a(\nu, \rho)$ |
| $\delta^2_\pi(\Theta_1, \Theta_2)$ | $\pi + (1-\pi)\rho^2$ | $\pi + (1-\pi)\delta^2_a(\nu, \rho)$ |

# Example: Test equality of two normal means

- $f(\mathbf{x}|\theta_1, \theta_2, \phi\Omega) = N\Big((\theta_1, \theta_2), \phi\Omega\Big)$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$
- Normal-gamma prior under $H_2$

$$P(\theta_1, \theta_2|\phi, H_2) = N\Big((m_1, m_2), h\phi\Omega\Big)$$
$$P(\phi) = \text{Ga}(\nu/2, \nu/2)$$

| | $P(\theta_1, \theta_2\|\phi)$ | $P(\theta_1, \theta_2)$ |
|---|---|---|
| Absolutely continuous part | Bivariate normal$(\rho)$ | Bivariate $t(\rho, \nu)$ |
| $\delta_a^2(\Theta_1, \Theta_2)$ | $\rho^2$ | $\delta_a^2(\nu, \rho)$ |
| $\delta_\pi^2(\Theta_1, \Theta_2)$ | $\pi + (1-\pi)\rho^2$ | $\pi + (1-\pi)\delta_a^2(\nu, \rho)$ |

- Notes:

$$\delta_a^2(\nu, \rho) = \rho^2 + (1-\rho^2)\delta^2(\nu, 0)$$
$$\delta_\pi^2(\Theta_1, \Theta_2) \geq \delta_\pi^2(\Theta_1, \Theta_2|\phi) \geq \pi$$

# Example: Test equality of two normal means

- $f(\mathbf{x}|\theta_1, \theta_2, \phi\Omega) = N\Big((\theta_1, \theta_2), \phi\Omega\Big)$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$
- Normal-gamma prior under $H_2$

$$P(\theta_1, \theta_2|\phi, H_2) = N\Big((m_1, m_2), h\phi\Omega\Big)$$
$$P(\phi) = \text{Ga}(\nu/2, \nu/2)$$

| | $P(\theta_1, \theta_2\|\phi)$ | $P(\theta_1, \theta_2)$ |
|---|---|---|
| Absolutely continuous part | Bivariate normal$(\rho)$ | Bivariate $t(\rho, \nu)$ |
| $\delta_a^2(\Theta_1, \Theta_2)$ | $\rho^2$ | $\delta_a^2(\nu, \rho)$ |
| $\delta_\pi^2(\Theta_1, \Theta_2)$ | $\pi + (1-\pi)\rho^2$ | $\pi + (1-\pi)\delta_a^2(\nu, \rho)$ |

- Notes:

$$\delta_a^2(\nu, \rho) = \rho^2 + (1-\rho^2)\delta^2(\nu, 0)$$
$$\delta_\pi^2(\Theta_1, \Theta_2) \geq \delta_\pi^2(\Theta_1, \Theta_2|\phi) \geq \pi$$

- Prior $P(\theta|H_1) = f_s(\theta)$
- Posterior dependence, replace $\pi$ with $\pi^*$

# Example: Test equality of two exponential parameters

- $f(x_{ji}|\theta_j) = \theta_j e^{-\theta_j x}, j = 1, 2$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$

# Example: Test equality of two exponential parameters

- $f(x_{ji}|\theta_j) = \theta_j e^{-\theta_j x}, j = 1, 2$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$
- Independent exponential prior under $H_2$ (Christensen et al. 2010)

$$P(\theta_1, \theta_2 | H_2) = e^{-\theta_1 - \theta_2}, \quad \theta_i | H_2, i =, 1, 2, \text{ independent}$$

  - In this case,
    - $\delta_a^2(\Theta_1, \Theta_2) = 0$
    - Dependence index $\delta_\pi^2(\Theta_1, \Theta_2) = \pi$

# Example: Test equality of two exponential parameters

- $f(x_{ji}|\theta_j) = \theta_j e^{-\theta_j x}, j = 1, 2$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$
- Independent exponential prior under $H_2$ (Christensen et al. 2010)

$$P(\theta_1, \theta_2|H_2) = e^{-\theta_1 - \theta_2}, \quad \theta_i|H_2, i =, 1, 2, \text{ independent}$$

  - In this case,
    - $\delta_a^2(\Theta_1, \Theta_2) = 0$
    - Dependence index $\delta_\pi^2(\Theta_1, \Theta_2) = \pi$
  - Posterior dependence
    - Two samples: $\mathbf{x}_j = (x_{j1}, \cdots, x_{jn_j}), \; j = 1, 2$
    - Prior $P(\theta|H_1) = e^{-\theta}$
    - Bayes factor

$$B_{21} = \frac{n_1! n_2! (S_1 + S_2 + 1)^{n+1}}{n!(S_1 + 1)^{n_1+1}(S_2 + 1)^{n_2+1}}, \qquad S_j = \sum_{i=1}^{n_j} x_{ji}, \; j = 1, 2$$

    - $\delta_{\pi^*}^2(\Theta_1, \Theta_2|\mathbf{x}_1, \mathbf{x}_2) = \pi^*$

# Example: Test equality of two exponential parameters

- $f(x_{ji}|\theta_j) = \theta_j e^{-\theta_j x}, j = 1, 2$
- $H_1 : \theta_1 = \theta_2 = \theta$ with $P(H_1) = \pi$, against $H_2 : \theta_1 \neq \theta_2$
- Independent exponential prior under $H_2$ (Christensen et al. 2010)

$$P(\theta_1, \theta_2|H_2) = e^{-\theta_1-\theta_2}, \quad \theta_i|H_2, i =, 1, 2, \text{ independent}$$

  - In this case,
    - $\delta_a^2(\Theta_1, \Theta_2) = 0$
    - Dependence index $\delta_\pi^2(\Theta_1, \Theta_2) = \pi$
  - Posterior dependence
    - Two samples: $\mathbf{x}_j = (x_{j1}, \cdots, x_{jn_j}), \ j = 1, 2$
    - Prior $P(\theta|H_1) = e^{-\theta}$
    - Bayes factor

$$B_{21} = \frac{n_1!n_2!(S_1 + S_2 + 1)^{n+1}}{n!(S_1 + 1)^{n_1+1}(S_2 + 1)^{n_2+1}}, \qquad S_j = \sum_{i=1}^{n_j} x_{ji}, \ j = 1, 2$$

    - $\delta_{\pi^*}^2(\Theta_1, \Theta_2|\mathbf{x}_1, \mathbf{x}_2) = \pi^*$
- Other singular bivariate exponential priors: Marshall-Olkin, bivariate autoregressive exponential, Gumbel and McKay with a singular part