FSAN/ELEG815: Statistical Learning

Gonzalo R. Arce

Department of Electrical and Computer Engineering
University of Delaware

I: Review of Probability

Signal Characterization

- Assumption: Many methods take $\{x(n)\}$ to be deterministic
- Reality: Real world signals are usually statistical in nature
  - Thus,

$$\ldots x(-1), x(0), x(1), \ldots$$

    can be interpreted as a sequence of random variables.
  - We begin by analyzing each observation $x(n)$ as a *R.V.*
  - Then, to capture dependencies, we consider random vectors

$$\ldots \underbrace{x(n), x(n+1), \ldots, x(n+N-1)}_{\underline{x}(n)}, x(n+N), \ldots$$

Random Variables

Definition

For a space S, the subsets, or events of S, have associated probabilities.

- To every event δ, we assign a number $x(\delta)$, which is called a *R.V.*
- The distribution function of x is

$$\Pr\{x \leq x_0\} = F_x(x_0) \qquad -\infty < x_0 < \infty$$

Properties:

1. $F(+\infty) = 1, \quad F(-\infty) = 0$
2. $F(x)$ is continuous from the right

$$F(x^+) = F(x)$$

3. $\Pr\{x_1 < x \leq x_2\} = F(x_2) - F(x_1)$

Example

Fair toss of two coins: H=heads, T=Tails

Define numerical assignments:

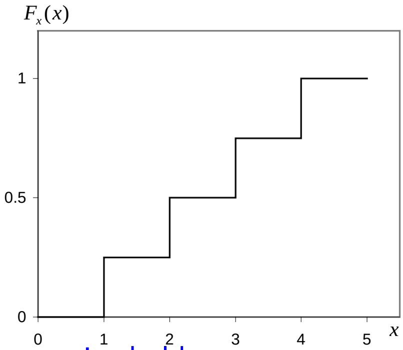
Events(δ)	Prob.	X(δ)	Y(δ)
HH	1/4	1	-100
HT	1/4	2	-100
TH	1/4	3	-100
TT	1/4	4	500

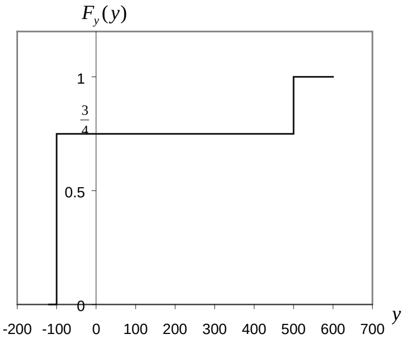
This assignments yield different distribution functions

$$
\begin{aligned}
F_x(2) &= \Pr\{HH, HT\} \\
F_y(2) &= \Pr\{HH, HT, TH\}
\end{aligned}
$$

How do we attain an intuitive interpretation of the distribution function?

Distribution Plots





Note properties hold:

1. $F(+\infty) = 1, \quad F(-\infty) = 0$
2. $F(x)$ is continuous from the right

$$F(x^+) = F(x)$$

3. $\Pr\{x_1 < x \leq x_2\} = F(x_2) - F(x_1)$

Definition

The probability density function is defined as,

$$f(x) = \frac{dF(x)}{dx}$$

$$\text{or} \quad F(x) = \int_{-\infty}^{x} f(x)dx$$

$$\text{Thus} \quad F(\infty) = 1 \Rightarrow \int_{-\infty}^{\infty} f(x)dx = 1$$

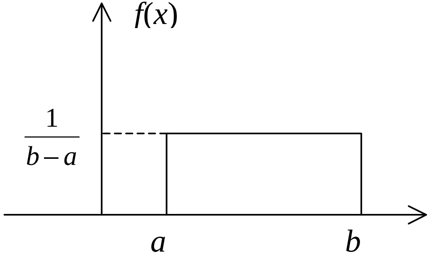
Types of distributions:

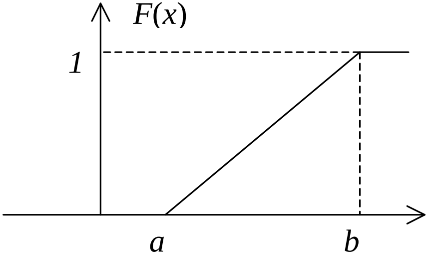
- Continuous: $\Pr\{x = x_0\} = 0 \quad \forall x_0$
- Discrete: $F(x_i) - F(x_i^-) = \Pr\{x = x_i\} = P_i$
  - In which case $f(x) = \sum_i P_i \delta(x - x_i)$
- Mixed: discontinuous but not discrete

Distribution examples

Uniform: $x \sim U(a,b) \quad a < b$

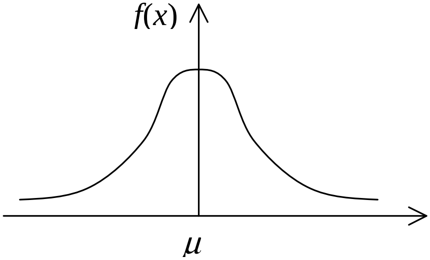
$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{else} \end{cases}$$

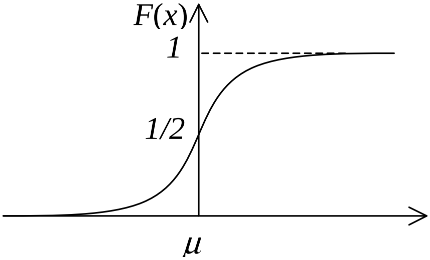




Gaussian: $x \sim N(\mu, \sigma)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

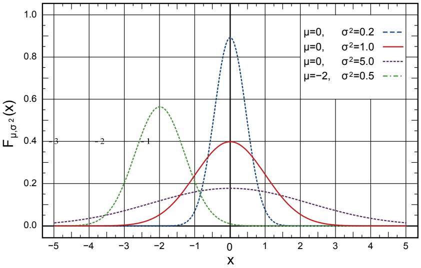


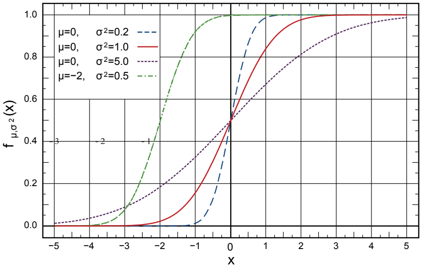


Gaussian Distribution Example

Example

Consider the Normal (Gaussian) distribution PDF and CDF for $\mu = 0, \sigma^2 = 0.2, 1.0, 5.0$ and $\mu = -2, \sigma^2 = 0.5$





Binomial: $x \sim B(p,q) \quad p+q=1$

Example

Toss a coin n times. What is the probability of getting k heads?

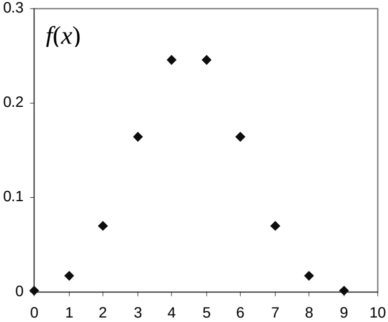
For $p+q=1$, where q is probability of a tail, and p is the probability of a head:

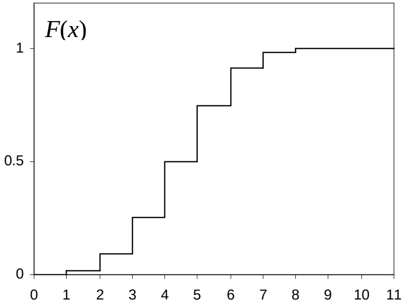
$$\begin{aligned}
\Pr\{x=k\} &= \binom{n}{k} p^k q^{n-k} \qquad \left[\text{NOTE: } \binom{n}{k} = \frac{n!}{(n-k)!k!}\right] \\
\Rightarrow f(x) &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta(x-k) \\
\Rightarrow F(x) &= \sum_{k=0}^{m} \binom{n}{k} p^k q^{n-k} \quad m \leq x < m+1
\end{aligned}$$

Binomial Distribution Example I

Example

Toss a coin n times. What is the probability of getting k heads? For $n = 9, p = q = \frac{1}{2}$ (fair coin)

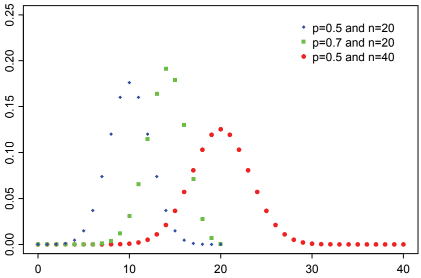


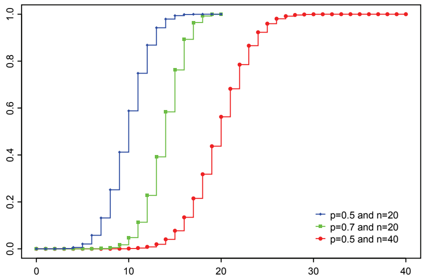


Binomial Distribution Example II

Example

Toss a coin n times. What is the probability of getting k heads? For $n = 20, p = 0.5, 0.7$ and $n = 40, p = 0.5$.





Conditional Distributions

Definition

The conditional distribution of x given event "M" has occurred is

$$
\begin{aligned}
F_x(x_0|M) &= \text{Pr}\{x \leq x_0|M\} \\
&= \frac{\text{Pr}\{x \leq x_0, M\}}{\text{Pr}\{M\}}
\end{aligned}
$$

Example

Suppose $M = \{x \leq a\}$, then

$$F_x(x_0|M) = \frac{\text{Pr}\{x \leq x_0, M\}}{\text{Pr}\{x \leq a\}}$$

If $x_0 \geq a$, what happens?

Special Cases

Special Case: $x_0 \geq a$

$$\Pr\{x \leq x_0, x \leq a\} = \Pr\{x \leq a\}$$

$$\Rightarrow F_x(x_0|M) = \frac{\Pr\{x \leq x_0, M\}}{\Pr\{x \leq a\}} = \frac{\Pr\{x \leq a\}}{\Pr\{x \leq a\}} = 1$$

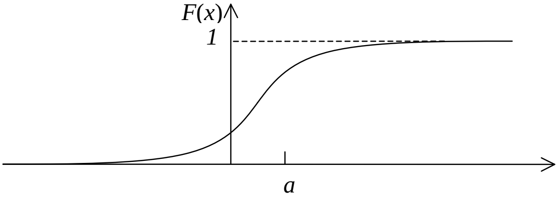
Special Case: $x_0 \leq a$

$$\begin{aligned} \Rightarrow F_x(x_0|M) &= \frac{\Pr\{x \leq x_0, M\}}{\Pr\{x \leq a\}} = \frac{\Pr\{x \leq x_0\}}{\Pr\{x \leq a\}} \\ &= \frac{F_x(x_0)}{F_x(a)} \end{aligned}$$

Conditional Distribution Example

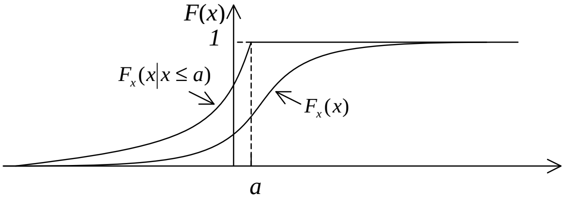
Example

Suppose



What does $F_x(x|M)$ look like? Note $M = \{x \leq a\}$.

$$\Rightarrow F_x(x_0|M) = \begin{cases} \frac{F_x(x_0)}{F_x(a)} & x \leq a \\ 1 & a \leq x \end{cases}$$



- Distribution properties hold for conditional cases:
  - Limiting cases: $F(\infty|M) = 1$ and $F(-\infty|M) = 0$
  - Probability range: $\Pr\{x_0 \leq x \leq x_1|M\} = F(x_1|M) - F(x_0|M)$
  - Density–distribution relations:

$$f(x|M) = \frac{\partial F(x|M)}{\partial x}$$

$$F(x_0|M) = \int_{-\infty}^{x_0} f(x|M)dx$$

Example (Fair Coin Toss)

Toss a fair coin 4 times. Let x be the number of heads. Determine $\Pr\{x=k\}$.

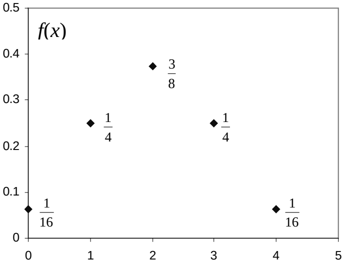
Recall

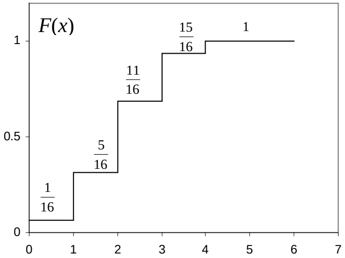
$$\Pr\{x=k\} = \binom{n}{k} p^k q^{n-k}$$

In this case

$$\begin{aligned}
\Pr\{x=k\} &= \binom{4}{k}\left(\frac{1}{2}\right)^4 \\
\Pr\{x=0\} &= \Pr\{x=4\} = \frac{1}{16} \\
\Pr\{x=1\} &= \Pr\{x=3\} = \frac{1}{4} \\
\Pr\{x=2\} &= \frac{3}{8}
\end{aligned}$$

Density and Distribution Plots for Fair Coin ($n = 4$) Ex.





What type of distribution is this? Discrete. Thus,

$$F(x_i) - F(x_i^-) = \mathrm{Pr}\{x = x_i\} = P_i$$

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \sum_i P_i \delta(x - x_i)dx$$

Conditional Case

Example (Conditional Fair Coin Toss)

Toss a fair coin 4 times. Let x be the number of heads. Suppose $M =$ [at least one flip produces a head]. Determine $\Pr\{x = k|M\}$.

Recall,

$$\Pr\{x = k|M\} = \frac{\Pr\{x = k, M\}}{\Pr\{M\}}$$

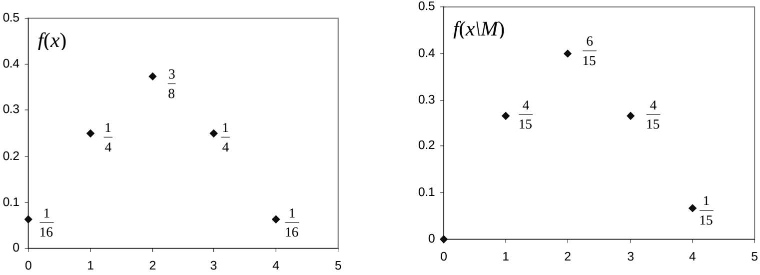
Thus first determine $\Pr\{M\}$

$$\begin{aligned}\Pr\{M\} &= 1 - \Pr\{\text{No heads}\} \\ &= 1 - \frac{1}{16} \\ &= \frac{15}{16}\end{aligned}$$

Next determine $\Pr\{x = k|M\}$ for the individual cases, $k = 0, 1, 2, 3, 4$

$$
\begin{aligned}
\Pr\{x = 0|M\} &= \frac{\Pr\{x = 0, M\}}{\Pr\{M\}} = 0 \\
\Pr\{x = 1|M\} &= \frac{\Pr\{x = 1, M\}}{\Pr\{M\}} \\
&= \frac{\Pr\{x = 1\}}{\Pr\{M\}} = \frac{1/4}{15/16} = \frac{4}{15} \\
\Pr\{x = 2|M\} &= \frac{\Pr\{x = 2\}}{\Pr\{M\}} = \frac{3/8}{15/16} = \frac{6}{15} \\
\Pr\{x = 3|M\} &= \frac{4}{15} \\
\Pr\{x = 4|M\} &= \frac{1}{15}
\end{aligned}
$$

Conditional and Unconditional Density Functions



Are they proper density functions?

Total Probability and Bayes' Theorem

Let $M_1, M_2, \ldots, M_n$ forms a partition of S, i.e.

$$\bigcup_i M_i = S \quad \text{and} \quad M_i \bigcap_{i \neq j} M_j = \phi$$

Then

$$\begin{aligned} F(x) &= \sum_i F_x(x|M_i)\text{Pr}(M_i) \\ f(x) &= \sum_i f_x(x|M_i)\text{Pr}(M_i) \end{aligned}$$

Aside

$$\text{Pr}\{A|B\} = \frac{\text{Pr}\{A,B\}}{\text{Pr}\{B\}} = \frac{\text{Pr}\{B,A\}\text{Pr}\{A\}}{\text{Pr}\{B\}\text{Pr}\{A\}} = \frac{\text{Pr}\{B|A\}\text{Pr}\{A\}}{\text{Pr}\{B\}}$$

Bayes' Theorem Example

Consider the following events and their probabilities:

A: patient has liver disease. $\Pr\{A\} = 0.1$.

B: patient is an alcoholic. $\Pr\{B\} = 0.05$.

Among all patients diagnosed with liver disease, 7% are alcoholics. $\Pr\{B|A\} = 0.07$.

Determine: patients' probability of having liver disease if they are an alcoholic, i.e. $\Pr\{A|B\}$.

Using the Bayes' Theorem:

$$\Pr\{A|B\} = \frac{\Pr\{B|A\}\Pr\{A\}}{\Pr\{B\}} = \frac{0.07 \times 0.1}{0.05} = 0.14$$

Mean, Median and variance

Definitions

$$\text{Mean} \quad E\{x\} = \int_{-\infty}^{\infty} x f(x) dx$$

$$\text{Conditional Mean} \quad E\{x|M\} = \int_{-\infty}^{\infty} x f(x|M) dx$$

Median

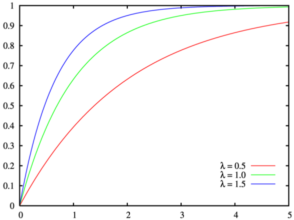
Definitions

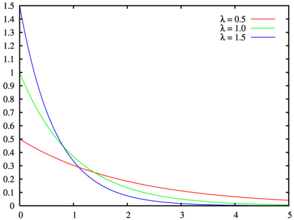
$$\text{Median} = m \implies \int_{-\infty}^{m} f(x)dx = \int_{m}^{\infty} f(x)dx = \frac{1}{2}$$

$$\text{Median} \implies Pr\{x \leq m\} = Pr\{x \geq m\}$$

Example

Let $x \sim \lambda \exp^{-\lambda x} U(x)$. Then $m = \frac{\ln(2)}{\lambda}$





Definition (Variance)

$$\text{Variance} \quad \sigma^2 = \int_{-\infty}^{\infty} (x-\eta)^2 f(x) dx$$

where $\eta = E\{x\}$. Thus,

$$\sigma^2 = E\{(x-\eta)^2\} = E\{x^2\} - E^2\{x\}$$

Example

For $x \sim N(\eta, \sigma^2)$, determine the variance.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\eta)^2}{2\sigma^2}}$$

Note: $f(x)$ is symmetric about $x = \eta \Rightarrow E\{x\} = \eta$

Also

$$\int_{-\infty}^{\infty} f(x) dx = 1 \Rightarrow \int_{-\infty}^{\infty} e^{-\frac{(x-\eta)^2}{2\sigma^2}} dx = \sqrt{2\pi}\sigma$$

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\eta)^2}{2\sigma^2}} dx = \sqrt{2\pi}\sigma$$

Differentiating w.r.t. σ:

$$\Rightarrow \int_{-\infty}^{\infty} \frac{(x-\eta)^2}{\sigma^3} e^{-\frac{(x-\eta)^2}{2\sigma^2}} dx = \sqrt{2\pi}$$

Rearranging yields

$$\int_{-\infty}^{\infty} (x-\eta)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\eta)^2}{2\sigma^2}} dx = \sigma^2$$

or

$$E\{(x-\eta)^2\} = \sigma^2$$

Definition (Moments)

- Moments

$$m_n = E\{x^n\} = \int_{-\infty}^{\infty} x^n f(x)dx$$

- Central Moments

$$\mu_n = E\{(x-\eta)^n\} = \int_{-\infty}^{\infty} (x-\eta)^n f(x)dx$$

From the binomial theorem

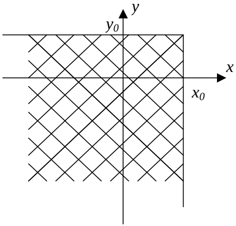
$$\begin{aligned}\mu_n &= E\{(x-\eta)^n\} = E\left\{\sum_{k=0}^{n}\binom{n}{k}x^k(-\eta)^{n-k}\right\} \\ &= \sum_{k=0}^{n}\binom{n}{k}m_k(-\eta)^{n-k}\end{aligned}$$

$$\Rightarrow \quad \mu_0 = 1, \quad \mu_1 = 0, \quad \mu_2 = \sigma^2, \quad \mu_3 = m_3 - 3\eta m_2 + 2\eta^3$$

Bivariate Statistics

Given two *RVs*, x and y, the bivariate (joint) distribution is given by

$$F(x_0, y_0) = \text{Pr}\{x \leq x_0, y \leq y_0\}$$

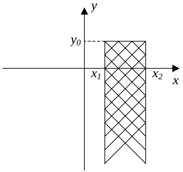


Properties:

- $F(-\infty, y) = F(x, -\infty) = 0$
- $F(\infty, \infty) = 1$
- $F_x(x) = F(x, \infty), \quad F_y(y) = F(\infty, y)$

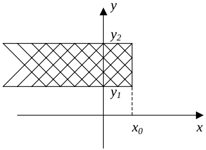
Special Cases

Case 1: $M = \{x_1 \leq x \leq x_2, y \leq y_0\}$



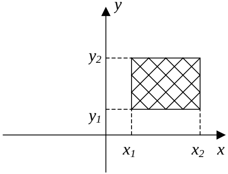
$$\Rightarrow \Pr\{M\} = F(x_2, y_0) - F(x_1, y_0)$$

Case 2: $M = \{x \leq x_0, y_1 \leq y \leq y_2\}$



$$\Rightarrow \Pr\{M\} = F(x_0, y_2) - F(x_0, y_1)$$

Case 3: $M = \{x_1 \leq x \leq x_2, y_1 \leq y \leq y_2\}$ Then



and

$$\Pr\{M\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + \underbrace{F(x_1, y_1)}_{\downarrow}$$

Added back because this region was subtracted twice

Definition (Joint Statistics)

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

and

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\alpha,\beta) d\alpha d\beta$$

In general, for some region M, the joint statistics are

$$\text{Pr}\{(x,y) \in M\} = \int\int_{M} f(x,y) dx dy$$

Marginal Statistics: $F_x(x) = F(x,\infty)$ and $F_y(y) = F(\infty,y)$

$$\Rightarrow f_x(x) = \int_{-\infty}^{\infty} f(x,y) dy$$

$$\Rightarrow f_y(y) = \int_{-\infty}^{\infty} f(x,y) dx$$

Independence

Definition (Independence)

Two *RVs* x and y are statistically independent if for arbitrary events (regions) $x \in A$ and $y \in B$,

$$\Pr\{x \in A, y \in B\} = \Pr\{x \in A\}\Pr\{y \in B\}$$

Letting $A = \{x \leq x_0\}$ and $B = \{y \leq y_0\}$, we see x and y are independent iff

$$F_{x,y}(x,y) = F_x(x)F_y(y)$$

and by differentiation

$$f_{x,y}(x,y) = f_x(x)f_y(y)$$

Joint Moments

For *RVs* x and y and function $z = g(x, y)$

$$\begin{aligned} E\{z\} &= \int_{-\infty}^{\infty} z f_z(z) dz \\ E\{g(x,y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy \end{aligned}$$

Definition (Covariance)

For *RVs* x and y,

$$\begin{aligned} C_{xy} &= \text{Cov}(x,y) \\ &= E[(x-\eta_x)(y-\eta_y)] \\ &= E[xy] - \eta_x E[y] - \eta_y E[x] + \eta_x\eta_y \\ &= E[xy] - \eta_x\eta_y \end{aligned}$$

Definition (Correlation Coefficient)

The correlation coefficient is given by

$$r = \frac{C_{xy}}{\sigma_x \sigma_y}$$

Note that

$$\begin{aligned} 0 &\leq E\{[a(x-\eta_x)+(y-\eta_y)]^2\} \\ &= E\{(x-\eta_x)^2\}a^2 + 2E\{(x-\eta_x)(y-\eta_y)\}a + E\{(y-\eta_y)^2\} \\ &= \sigma_x^2 a^2 + 2C_{xy}a + \sigma_y^2 \end{aligned}$$

This is a positive quadratic function of a

$\Rightarrow$ Roots are imaginary and discriminant is non-positive

$$\begin{aligned} \sqrt{4C_{xy}^2 - 4\sigma_x^2\sigma_y^2} &\rightarrow \text{imaginary} \\ \Rightarrow 4C_{xy}^2 - 4\sigma_x^2\sigma_y^2 &\leq 0 \\ \Rightarrow C_{xy}^2 &\leq \sigma_x^2\sigma_y^2 \end{aligned}$$

Thus,

$$|C_{xy}| \leq \sigma_x\sigma_y \quad \text{and} \quad |r| = \frac{|C_{xy}|}{\sigma_x\sigma_y} \leq 1$$

Definition (Uncorrelated)

Two *RVs* are uncorrelated if their covariance is zero

$$\begin{aligned} C_{xy} &= 0 \\ \Rightarrow r &= \frac{C_{xy}}{\sigma_x\sigma_y} = 0 \\ &= \frac{E\{xy\} - E\{x\}E\{y\}}{\sigma_x\sigma_y} = 0 \\ \Rightarrow E\{xy\} &= E\{x\}E\{y\} \end{aligned}$$

Thus

$$C_{xy} = 0 \Leftrightarrow E\{xy\} = E\{x\}E\{y\}$$

Result

If x and y are independent, then

$$E\{xy\} = E\{x\}E\{y\}$$

and x and y are uncorrelated

Note: Converse is not true (in general)

- Converse only holds for Gaussian *RVs*
- Independence is a stronger condition than uncorrelated

Definition (Orthogonality)

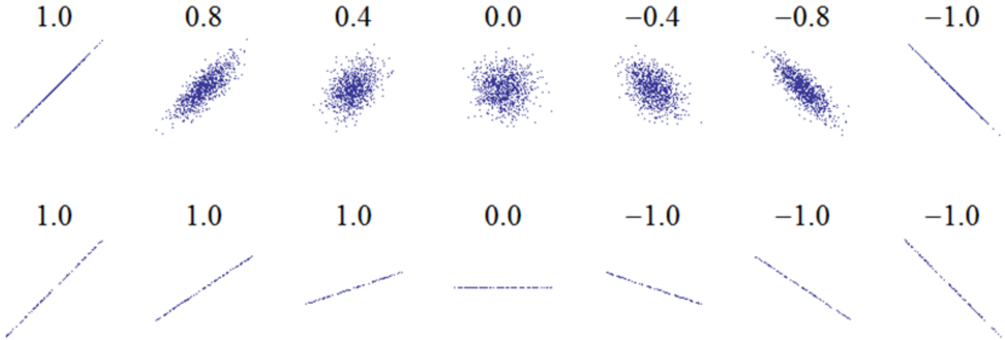
Two *RVs* are orthogonal if

$$E\{xy\} = 0$$

Note: If x and y are correlated, they are not orthogonal

Example

Consider the correlation between two *RV*'s, x and y, with samples shown in a scatter plot



Sequences and Vectors of Random Variables

Definition (Vector Distribution)

Let $\{x\}$ be a sequence of *RVs*. Take N samples to form the random vector

$$\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$$

Then the vector distribution function is

$$\begin{aligned} F_{\mathbf{x}}(\mathbf{x}^0) &= \Pr\{x_1 \leq x_1^0, x_2 \leq x_2^0, \ldots, x_N \leq x_N^0\} \\ &\triangleq \Pr\{\mathbf{x} \leq \mathbf{x}^0\} \end{aligned}$$

The density function is given by

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\partial^N F_{\mathbf{x}(\mathbf{x})}}{\partial x_1 \partial x_2 \ldots \partial x_N}$$

$$F_{\mathbf{x}}(\mathbf{x}^0) = \int_{-\infty}^{\mathbf{x}_1^0} \int_{-\infty}^{\mathbf{x}_2^0} \cdots \int_{-\infty}^{\mathbf{x}_N^0} f_{\mathbf{x}}(\mathbf{x}) dx_1 dx_2 \ldots dx_N$$

Properties:

$$
\begin{aligned}
F_{\mathbf{x}}([\infty, \infty, \cdots, \infty]^T) &= 1 \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} &= 1 \\
F_{\mathbf{x}}([x_1, x_2, \cdots, -\infty, \cdots, x_N]^T) &= 0
\end{aligned}
$$

Also

$$
\begin{aligned}
F([\infty, x_2, x_3, \cdots, x_N]^T) &= F([x_2, x_3, \cdots, x_N]^T) \\
\int_{-\infty}^{\infty} f([x_1, x_2, x_3, \cdots, x_N]^T) dx_1 &= f([x_2, x_3, \cdots, x_N]^T)
\end{aligned}
$$

- Setting $x_i = \infty$ in the cdf eliminates this sample
- Integrating over $(-\infty, \infty)$ along x_i in the pdf eliminates this sample

Joint Distribution

Definitions (Joint Distribution and Density)

Given two random vectors $\mathbf{x}$ and $\mathbf{y}$, the joint distribution and density are

$$
\begin{aligned}
F_{\mathbf{xy}}(\mathbf{x}^0, \mathbf{y}^0) &= \Pr\{\mathbf{x} \leq \mathbf{x}^0, \mathbf{y} \leq \mathbf{y}^0\} \\
f_{\mathbf{xy}}(\mathbf{x}, \mathbf{y}) &= \frac{\partial^N \partial^M F_{\mathbf{xy}}(\mathbf{x}, \mathbf{y})}{\partial x_1 \partial x_2 \cdots \partial x_N \partial y_1 \partial y_2 \cdots \partial y_M}
\end{aligned}
$$

Definition (Vector Independence)

The vectors are independent iff

$$F_{\mathbf{xy}}(\mathbf{x}, \mathbf{y}) = F_{\mathbf{x}}(\mathbf{x}) F_{\mathbf{y}}(\mathbf{y})$$

or equivalently

$$f_{\mathbf{xy}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{x}}(\mathbf{x}) f_{\mathbf{y}}(\mathbf{y})$$

Expectations & Moments

Objective: Obtain partial description of process generating $\mathbf{x}$

Solution: Use moments

The first moment, or mean, is

$$\begin{aligned}\mathbf{m}_{\mathbf{x}} = E\{\mathbf{x}\} &= [m_1, m_2, \ldots, m_N]^T \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathbf{x} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}\end{aligned}$$

$$\begin{aligned}\Rightarrow m_k &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_k f_{\mathbf{x}}(\mathbf{x}) dx_1 dx_2 \cdots dx_N \\ &= \int_{-\infty}^{\infty} x_k \underbrace{f_{x_k}(x_k)} dx_k\end{aligned}$$

$\Uparrow$ marginal distribution of x_k

Definition (Correlation Matrix)

A complete set of second moments is given by the correlation matrix

$$\begin{aligned}\mathbf{R_x} &= E\{\mathbf{x}\mathbf{x}^H\} = E\{\mathbf{x}\mathbf{x}^{*^T}\} \\ &= \begin{bmatrix} E\{|x_1|^2\} & E\{x_1x_2^*\} & \cdots & E\{x_1x_N^*\} \\ E\{x_2x_1^*\} & E\{|x_2|^2\} & \cdots & E\{x_2x_N^*\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{x_Nx_1^*\} & E\{x_Nx_2^*\} & \cdots & E\{|x_N|^2\} \end{bmatrix}\end{aligned}$$

Result

The correlation matrix is Hermitian symmetric

$$\begin{aligned}(\mathbf{R_x})^H &= (E\{\mathbf{x}\mathbf{x}^H\})^H \\ &= E\{(\mathbf{x}\mathbf{x}^H)^H\} \\ &= E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{R_x}\end{aligned}$$

Definition (Covariance Matrix)

The set of second central moments is given by the covariance

$$
\begin{aligned}
\mathbf{C_x} &= E\{(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^H\} \\
&= E\{\mathbf{x}\mathbf{x}^H\} - \mathbf{m_x}E\{x^H\} - E\{\mathbf{x}\}\mathbf{m_x}{}^H + \mathbf{m_x}\mathbf{m_x}{}^H \\
&= \mathbf{R_x} - \mathbf{m_x}\mathbf{m_x}{}^H
\end{aligned}
$$

Result

The covariance is Hermitian symmetric

$$\mathbf{C_x} = \mathbf{C_x}{}^H$$

Result

The correlation and covariance matrices are positive semi-definite

$$\mathbf{a}^H \mathbf{R_x} \mathbf{a} \geq 0 \quad \mathbf{a}^H \mathbf{C_x} \mathbf{a} \geq 0 \quad (\forall \mathbf{a})$$

To prove this, note

$$\begin{aligned} \mathbf{a}^H \mathbf{R_x} \mathbf{a} &= \mathbf{a}^H E\{\mathbf{x}\mathbf{x}^H\}\mathbf{a} \\ &= E\{\mathbf{a}^H \mathbf{x}\mathbf{x}^H \mathbf{a}\} \\ &= E\{(\mathbf{a}^H \mathbf{x})(\mathbf{a}^H \mathbf{x})^H\} \\ &= E\{|\mathbf{a}^H \mathbf{x}|^2\} \geq 0 \end{aligned}$$

For most cases, $\mathbf{R}$ and $\mathbf{C}$ are positive define

$$\mathbf{a}^H \mathbf{R_x} \mathbf{a} > 0 \quad \mathbf{a}^H \mathbf{C_x} \mathbf{a} > 0$$

$\Rightarrow$ no linear dependencies in $\mathbf{R_x}$ or $\mathbf{C_x}$

Definitions (Cross-Correlation and Cross-Covariance)

For random vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\begin{aligned}\text{Cross-correlation} \triangleq \mathbf{R}_{\mathbf{xy}} &= E\{\mathbf{x}\mathbf{y}^H\}\\ \text{Cross-covariance} \triangleq \mathbf{C}_{\mathbf{xy}} &= E\{(\mathbf{x}-\mathbf{m}_{\mathbf{x}})(\mathbf{y}-\mathbf{m}_{\mathbf{y}})^H\}\\ &= \mathbf{R}_{\mathbf{xy}} - \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{y}}^H\end{aligned}$$

Definition (Uncorrelated Vectors)

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated if

$$\mathbf{C}_{\mathbf{xy}} = \mathbf{R}_{\mathbf{xy}} - \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{y}}^H = 0$$

or equivalently

$$\mathbf{R}_{\mathbf{xy}} = E\{\mathbf{x}\mathbf{y}^H\} = \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{y}}^H$$

Note that as in the scalar case

$$\begin{aligned} \text{independence} &\Rightarrow \text{uncorrelated} \\ \text{uncorrelated} &\not\Rightarrow \text{independence} \end{aligned}$$

Also, $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if

$$\mathbf{R}_{\mathbf{xy}} = E\{\mathbf{x}\mathbf{y}^H\} = \mathbf{0}$$

Example

Let $\mathbf{x}$ and $\mathbf{y}$ be the same dimension. If

$$\mathbf{z} = \mathbf{x} + \mathbf{y}$$

find $\mathbf{R}_{\mathbf{z}}$ and $\mathbf{C}_{\mathbf{z}}$

By definition

$$\begin{aligned}\mathbf{R_z} &= E\{(\mathbf{x}+\mathbf{y})(\mathbf{x}+\mathbf{y})^H\} \\ &= E\{\mathbf{x}\mathbf{x}^H\}+E\{\mathbf{x}\mathbf{y}^H\}+E\{\mathbf{y}\mathbf{x}^H\}+E\{\mathbf{y}\mathbf{y}^H\} \\ &= \mathbf{R_x}+\mathbf{R_{xy}}+\mathbf{R_{yx}}+\mathbf{R_y}\end{aligned}$$

Similarly

$$\mathbf{C_z} = \mathbf{C_x}+\mathbf{C_{xy}}+\mathbf{C_{yx}}+\mathbf{C_y}$$

Note: If $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated,

$$\mathbf{R_z} = \mathbf{R_x}+\mathbf{m_x}\mathbf{m_y}^H+\mathbf{m_y}\mathbf{m_x}^H+\mathbf{R_y}$$

and

$$\mathbf{C_z} = \mathbf{C_x}+\mathbf{C_y}$$

Tchebycheff Inequality

For any $\epsilon > 0$,

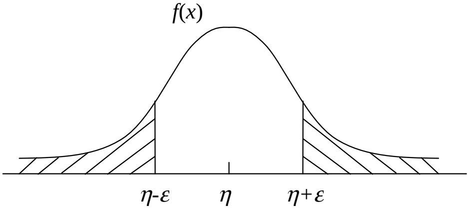
$$\Pr(|x-\eta| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

To prove this, note

$$\begin{aligned}\Pr(|x-\eta| \geq \epsilon) &= \int_{-\infty}^{\eta-\epsilon} f(x)dx + \int_{\eta+\epsilon}^{\infty} f(x)dx \\ &= \int_{|x-\eta|\geq\epsilon} f(x)dx\end{aligned}$$

Also note that

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{\infty} (x-\eta)^2 f(x)dx \\ &\geq \int_{|x-\eta|\geq\epsilon} (x-\eta)^2 f(x)dx\end{aligned}$$



$$\sigma^2 \geq \int\limits_{|x-\eta|\geq\epsilon} (x-\eta)^2 f(x)dx$$

Using the fact that $|x-\eta| \geq \epsilon$ in the above gives

$$\begin{aligned}\sigma^2 &\geq \epsilon^2 \int\limits_{|x-\eta|\geq\epsilon} f(x)dx \\ &= \epsilon^2 \text{Pr}\{|x-\eta| \geq \epsilon\}\end{aligned}$$

Rearranging gives the desired result

$$\Rightarrow \text{Pr}\{|x-\eta| \geq \epsilon\} \leq \left(\frac{\sigma}{\epsilon}\right)^2$$

QED

Markov's Inequality

If x is a non-negative RV, then for all $a > 0$

$$\Pr\{x \geq a\} \leq \frac{E\{x\}}{a}.$$

Proof:

$$\begin{aligned}
\Pr\{x \geq a\} &= \int_a^\infty f(x)dx \\
&\leq \int_a^\infty \frac{x}{a} f(x)dx \qquad \text{since } x \geq a \\
&\leq \frac{1}{a}\int_0^\infty x f(x)dx \\
&= \frac{E\{x\}}{a}.
\end{aligned}$$

Chernoff's Bounding Method

Let x be a RV on $\mathbb{R}$. Then for all $\epsilon > 0$

$$\Pr\{x \geq \epsilon\} \leq \min_{s>0} \quad e^{-s\epsilon} E\{e^{sx}\}.$$

To prove this for any $s > 0$:

$$\begin{aligned} \Pr\{x \geq \epsilon\} &= \Pr\{sx \geq s\epsilon\} \\ &= \Pr\{e^{sx} \geq e^{s\epsilon}\} \end{aligned}$$

Using Markov's Inequality:

$$\begin{aligned} \Pr\{x \geq \epsilon\} = \Pr\{e^{sx} \geq e^{s\epsilon}\} &\leq \frac{E\{e^{sx}\}}{e^{s\epsilon}} \\ &= e^{-s\epsilon} E\{e^{sx}\}. \end{aligned}$$

Hoeffding's Inequality

Consider $S_N = \sum_{i=1}^{N} x_i$ where $x_1, \ldots, x_N$ are independent RV's on $\mathbb{R}$ such that $a_i \leq x_i \leq b_i$. Then, for any $\epsilon > 0$

$$\Pr\{|S_N - E\{S_N\}| \geq \epsilon\} \quad \leq \quad 2e^{-2\epsilon^2/\sum(b_i - a_i)^2}$$

Lemma 1:
Let x be a random variable on $\mathbb{R}$ with $E\{x\} = 0$ and $a \leq x \leq b$. Then, for all $s > 0$

$$\boxed{E\{e^{sx}\} \leq e^{s^2(b-a)^2/8}} \qquad (*)$$

Hoeffding's Inequality

Apply Chernoff's bounding method i.e.:

$$\Pr\{x \geq \epsilon\} \leq \min_{s>0} \quad e^{-s\epsilon}E\{e^{sx}\})$$

to the random variable: $S_N - E\{S_N\}$,

$$\begin{aligned}
\Pr\{S_N - E\{S_N\} \geq \epsilon\} &\leq \min_{s>0} \quad e^{-s\epsilon}E\left\{e^{s(S_N - E\{S_N\})}\right\} \\
&\leq \min_{s>0} \quad e^{-s\epsilon}E\left\{e^{s\left(\sum_{i=1}^{N}(x_i - E\{x_i\})\right)}\right\} \\
\text{since the } x_i \text{ are independent} \quad &\leq \min_{s>0} \quad e^{-s\epsilon}\prod_{i=1}^{N}E\{e^{s(x_i - E\{x_i\})}\}
\end{aligned}$$

Applying **Lemma 1** $(*)$ to RV $y_i = x_i - E\{x_i\}$ where $E\{y_i\} = 0$:

$$E\{e^{s(x_i - E\{x_i\})}\} \leq e^{s^2(b_i - a_i)^2/8}$$

Hoeffding's Inequality

Substitute $E\{e^{s(x_i - E\{x_i\})}\} \leq e^{s^2(b_i-a_i)^2/8}$ in the previous Chernoff's bound:

$$\Pr\{S_N - E\{S_N\} \geq \epsilon\} \leq \min_{s>0} \quad e^{-s\epsilon} \prod_{i=1}^{N} E\{e^{s(x_i - E\{x_i\})}\}$$

we get:

$$\begin{aligned}\Pr\{S_N - E\{S_N\} \geq \epsilon\} &\leq \min_{s>0} e^{-s\epsilon} \prod_{i=1}^{N} e^{s^2(b_i-a_i)^2/8} \\ &= \min_{s>0} e^{-s\epsilon + \sum_{i=1}^{N}(s^2/8)(b_i-a_i)^2}\end{aligned}$$

It can be shown that the minimum is at $s = 4\epsilon/\sum(b_i - a_i)^2$.

Hoeffding's Inequality

$$\Pr\{S_N - E\{S_N\} \geq \epsilon\} \leq e^{-s\epsilon + \sum_{i=1}^{N}(s^2/8)(b_i - a_i)^2}$$

Substituting the minimum $(s = 4\epsilon / \sum_{i=1}^{N}(b_i - a_i)^2)$:

$$\Pr\{S_N - E\{S_N\} \geq \epsilon\} \quad \leq \quad e^{-2\epsilon^2 / \sum_{i=1}^{N}(b_i - a_i)^2}$$

If we consider $-x_1, \ldots, -x_N$ instead, we obtain:

$$\Pr\{S_N - E\{S_N\} \leq -\epsilon\} \leq e^{-2\epsilon^2 / \sum_{i=1}^{N}(b_i - a_i)^2}$$

By combining the two bounds, we finish the proof:

$$\Pr\{|S_N - E\{S_N\}| \geq \epsilon\} \quad \leq \quad 2e^{-2\epsilon^2 / \sum_{i=1}^{N}(b_i - a_i)^2}$$

Hoeffding's Inequality

Example:

Find the Hoeffding's Inequality of a random variable $x_i \overset{\text{i.i.d.}}{\sim}$ Ber(p).

Solution:

Consider the Hoeffding's Inequality:

$$\Pr\{|S_N - E\{S_N\}| \geq \epsilon\} \leq 2e^{-2\epsilon^2/\sum_{i=1}^N (b_i - a_i)^2}$$

Since $x_i \sim$ Ber(p), then $a_i = 0$, $b_i = 1$, $S_N = \sum_{i=1}^N x_i \sim Bin(N,p)$, and $E\{S_N\} = Np$. Taking $\epsilon = N\delta$ and applying Hoeffding's Inequality:

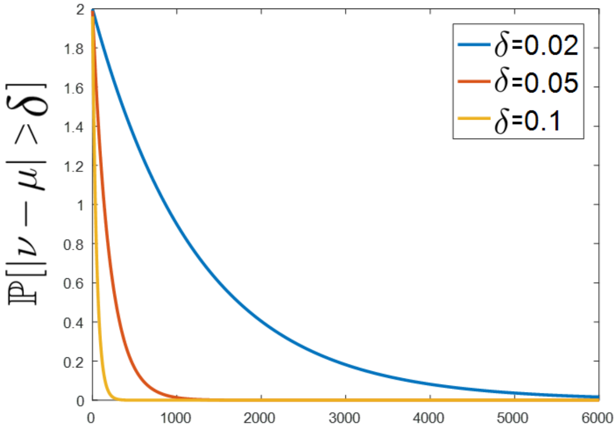
$$\Pr\left\{\left|\sum_{i=1}^N x_i - Np\right| \geq N\delta\right\} \leq 2e^{-2(N\delta)^2/\sum_{i=1}^N (1-0)^2}$$

$$\Pr\left\{\left|\frac{1}{N}\sum_{i=1}^N x_i - p\right| \geq \delta\right\} \leq 2e^{-2N\delta^2}$$

Hoeffding's Inequality

$$\Pr\left\{\left|\frac{1}{N}\sum_{i=1}^{N} x_i - p\right| \geq \delta\right\} \leq 2e^{-2N\delta^2}$$

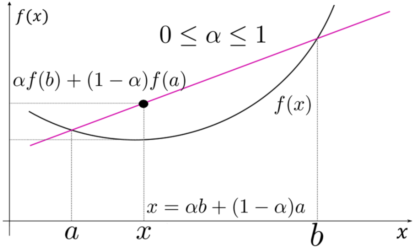
$$\Pr\{|\nu - \mu| \geq \delta\} \leq 2e^{-2N\delta^2}$$



Proof of Lemma 1:

Show that if $E\{x\}=0$ then $E\{e^{sx}\} \le e^{s^2(b-a)^2/8}$ for all $s>0$.
If $x \in [a,b]$ then the convexity of the function $f(x)=e^{sx}$ implies that

$$\begin{aligned} e^{sx} &\le \alpha f(b) + (1-\alpha) f(a), \\ e^{sx} &\le \alpha e^{sb} + (1-\alpha) e^{sa}, \text{ since } \alpha = \frac{x-a}{b-a}, \\ e^{sx} &\le \frac{x-a}{b-a} e^{sb} + \frac{b-x}{b-a} e^{sa} \end{aligned}$$



Proof of Lemma 1:

$$e^{sx} \leq \frac{x-a}{b-a}e^{sb} + \frac{b-x}{b-a}e^{sa}$$

Using the fact that $E\{x\} = 0$ we obtain:

$$\begin{aligned} E\{e^{sx}\} &\leq \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb}, \\ &= e^{sa}\left(\frac{b}{b-a} - \frac{a}{b-a}e^{s(b-a)}\right), \text{ since } y = e^{\ln(y)} \\ &= e^{\ln\left[e^{sa}\left(\frac{b-ae^{s(b-a)}}{b-a}\right)\right]} \end{aligned}$$

Thus,

$$E\{e^{sx}\} \leq e^{g(s)}$$

where $g(s) = sa + \ln(b - ae^{s(b-a)}) - \ln(b-a)$.

Proof of Lemma 1:

$$g(s) = sa + \ln(b - ae^{s(b-a)}) - \ln(b-a)$$

By Taylor's theorem:

$$g(s) = g(0) + g'(0)s + \frac{1}{2!}g''(\xi)s^2, \quad 0 \le \xi \le s$$

$$g(0) = 0, \qquad g'(0) = 0, \qquad g''(\xi) \le \frac{(b-a)^2}{4}$$

Substituting, we get: $\quad g(s) \le \frac{s^2(b-a)^2}{8}$.

Substituting in previous demonstration (i.e. $E\{e^{sx}\} \le e^{g(s)}$) :

$$\Longrightarrow \boxed{E\{e^{sx}\} \le e^{s^2(b-a)^2/8}} \qquad (*)$$