Final Project Timeline, Limit Theorem & Hazard Rate Examples

ECE 313

Probability with Engineering Applications

Lecture 23

Professor Ravi K. Iyer

Dept. of Electrical and Computer Engineering

University of Illinois at Urbana Champaign

Today's Topics

- **Final Project Timeline**

- **Review and Examples:**
  - Treatment of Failure Data and Hazard rates
  - Variance, Covariance, and Correlation
  - Central Limit Theorem

Announcements

- **Quiz 2, Next Tuesday, Nov. 19**
  - Hazard and failure rates
  - Topics covered in:
    - Lectures 20 and 23
    - Homework 8

- **Project Summary,** due **Saturday, Nov. 16, 5:00pm**
  - Send a one-paragraph summary of your goals and plans

Final Project Timeline

- **Project Summary: Saturday, Nov. 16, 5:00pm**
  - Submit a one-paragraph summary of your goals and overall plans
- **Short Progress Reports: Every Tuesday, starting Nov. 19.**
  - A short paragraph (3-4 lines) on what progress you made and what is the plan for the next week

- **First Presentation, next Thursday, Nov. 21, in the class**
  - 2 slides showing your overall plan, timeline, the division of tasks, and techniques are your going to use.
- **Intermediate Presentation, Tuesday, Dec. 3, in the class**
  - Reporting on your progress

- **Final Presentation, Saturday, Dec. 14, in the class**
- **Final Exam Review Session, Monday, Dec. 16, Time: TBD**
- **Final Project Report, December 19, 5:00pm.**

First Presentation

- Next Thursday, Nov. 21, 11:00am – 12:20pm, in the class
- Presentation: 5-6 mins, Questions: 2-3 mins => 2-3 slides

- Graded by the instructor, the TA, and the other students
- **Grading Scheme:**
  - **Presentation**
  - **Idea:**
    - What techniques or concepts from the class?
  - **Plan:**
    - Timeline
    - Division of tasks
  - **Content:**
    - Show a sample of your data with description
    - Show one or two graphs/histograms on your data & some insights
    - Show the data collection techniques and tools your are using.

Treatment of Failure Data Review

- Part failure data generally obtained from two sources: the failure times of various items in a population placed on a life test, or repair reports listing operating hours of replaced parts in equipment already in field use.
- Compute and plot either the failure density function or the instantaneous failure rate as a function of time.
- The data: a sequence of times to failure, but the failure density function and the hazard introduced as continuous variables.
- Compute a piecewise-continuous failure density function and hazard rate from the data.
- This is, a specific approach to the very general engineering problem of how to model a problem from certain qualitative knowledge about the system supported by quantitative data.

Treatment of Failure Data (Cont'd)

- Define piecewise-continuous failure density and hazard-rate in terms of the data.
- Assume that our data describe a set of N items placed in operation at time t=0. As time progresses, items fail, and at any time t the number of survivors in n(t).
- The empirical probability density function defined over the time interval $t_i < t \le t_i + \Delta t_i$, is given by the ratio of the number of failures occurring in the interval to the *size of the original population*, divided by the length of the time interval:

$$f_d(t) = \frac{[n(t_i) - n(t_i + \Delta t_i)]/N}{\Delta t_i} \; for \; t_i < t \le t_i + \Delta t_i$$

Treatment of Failure Data (Cont'd)

- The data hazard (inst. failure rate) over the interval $t_i < t \leq t_i + \Delta t_i$ is the ratio of the number of failures occurring in the time interval to the *number of survivors at the beginning of the time interval*, divided by the length of the time interval:

$$z_d(t) = \frac{[n(t_i) - n(t_i + \Delta t_i)] / n(t_i)}{\Delta t_i} \; for \; t_i < t \leq t_i + \Delta t_i$$

- Observation: the failure density function $f_d(t)$ is a measure of the *overall speed* at which failures are occurring, whereas the hazard rate $z_d(t)$ is a measure of the *instantaneous speed* of failure.
- Note: both $f_d(t)$ and $z_d(t)$ have the dimensions of inverse time (generally the time unit is hours).
- The choice of t_i and Δt_i in the above equations is unspecified and is best discussed in terms of the examples that follow.

Hazard Rate Example

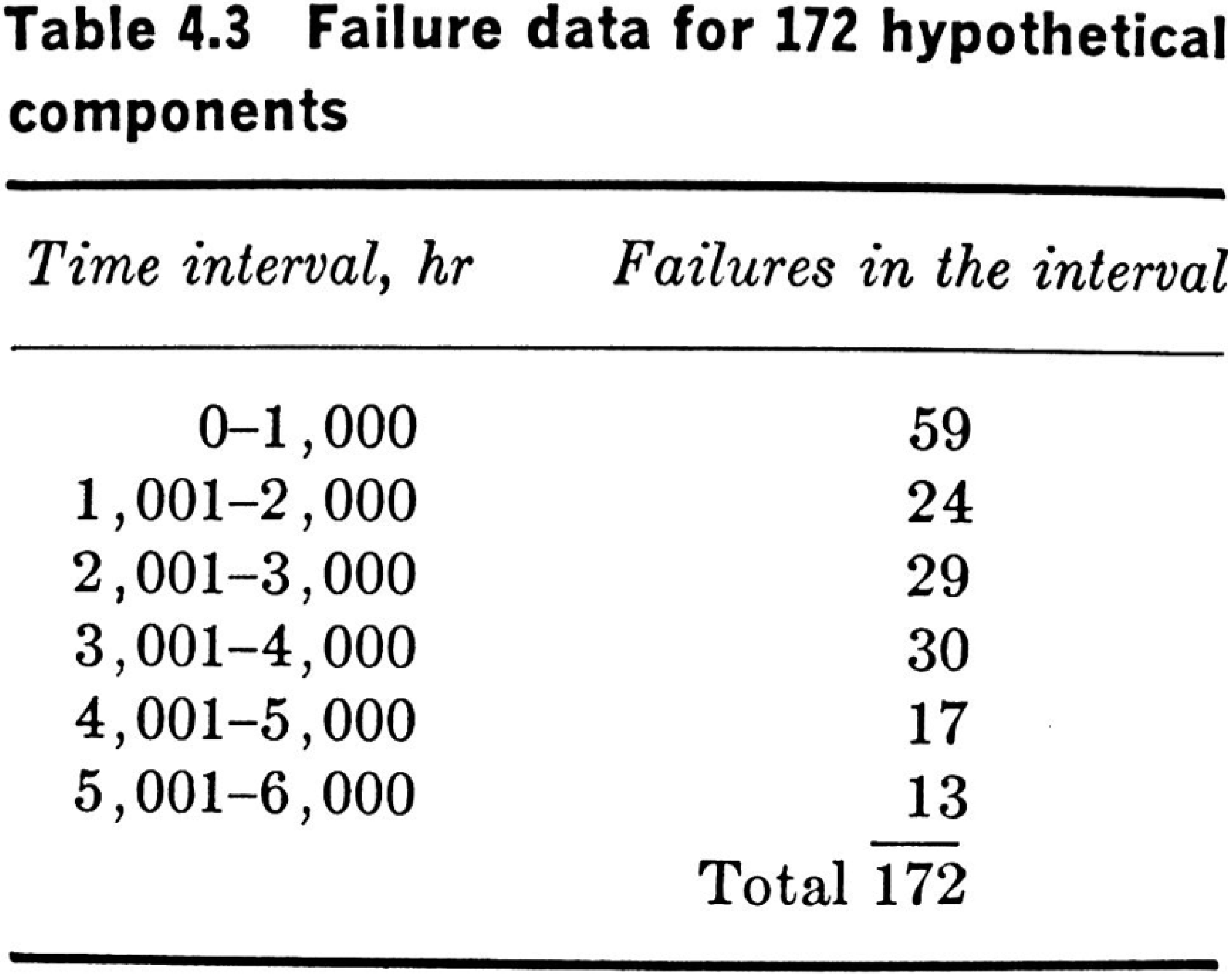
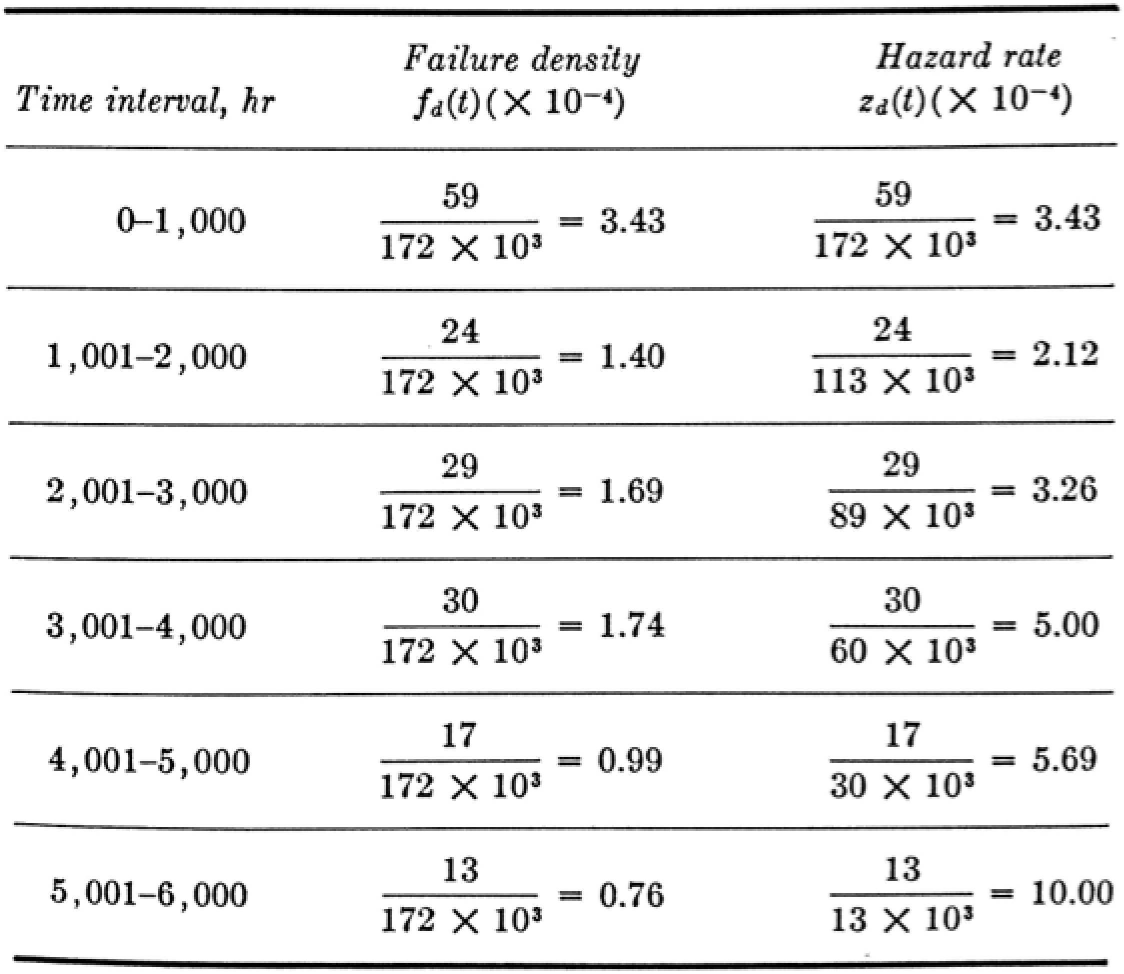


Table 4.3 Failure data for 172 hypothetical components

Time interval, hr	*Failures in the interval*
0–1,000	59
1,001–2,000	24
2,001–3,000	29
3,001–4,000	30
4,001–5,000	17
5,001–6,000	13
	Total 172

Hazard Example (Cont'd)

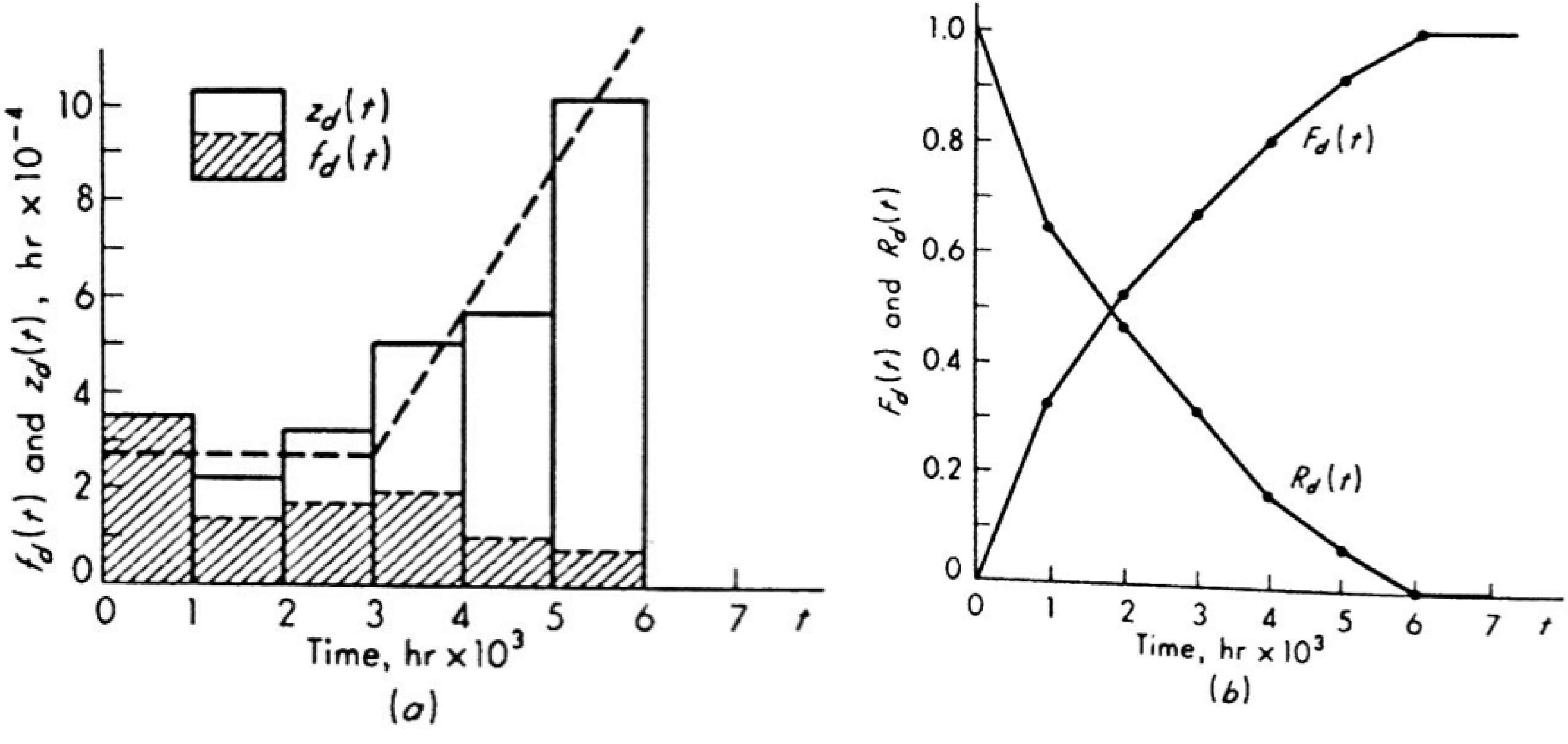
Table 4.4 Failure rates of hypothetical component



Time interval, hr	*Failure density* $f_d(t)(\times 10^{-4})$	*Hazard rate* $z_d(t)(\times 10^{-4})$
0–1,000	$\frac{59}{172 \times 10^3} = 3.43$	$\frac{59}{172 \times 10^3} = 3.43$
1,001–2,000	$\frac{24}{172 \times 10^3} = 1.40$	$\frac{24}{113 \times 10^3} = 2.12$
2,001–3,000	$\frac{29}{172 \times 10^3} = 1.69$	$\frac{29}{89 \times 10^3} = 3.26$
3,001–4,000	$\frac{30}{172 \times 10^3} = 1.74$	$\frac{30}{60 \times 10^3} = 5.00$
4,001–5,000	$\frac{17}{172 \times 10^3} = 0.99$	$\frac{17}{30 \times 10^3} = 5.69$
5,001–6,000	$\frac{13}{172 \times 10^3} = 0.76$	$\frac{13}{13 \times 10^3} = 10.00$

Hazard Rate Example (Cont'd)

Reliability functions: (a) $f_d(t)$ and $z_d(t)$; (b) $F_d(t)$ and $R_d(t)$



Covariance and Variance Review

- The covariance of any two random variables, *X* and *Y*, denoted by *Cov(X,Y)*, is defined by

$$\begin{aligned} Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \\ &= E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\ &= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- If X and Y are independent then it follows that $Cov(X,Y)=0$

- For any random variable X, Y, Z, and constant c, we have:
  1. Cov (X,X) = Var(X),
  2. Cov (X,Y) = Cov(Y,X),
  3. Cov (cX,Y) = cCov(X,y),
  4. Cof (X,Y+Z) = Cov(X,Y) + Cov(X,Z).

Covariance and Variance Review

- **Covariance and Variance of Sums of Random Variables**

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

- A useful expression for the variance of the sum of random variables can be obtained from the preceding equation

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

- If $X_i, i = 1,\ldots,n$ are independent random variables, then the above equation reduces to

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Covariance and Variance Example 1

- In general it can be shown that a positive value of $Cov(X,Y)$ is an indication that Y tends to increase as X does, whereas a negative value indicates that Y tends to decrease as X increases.
- **Example:** The joint density function of X,Y is:

$$f(x,y) = \frac{1}{y}e^{-(y+x/y)},\ 0 < x, y < \infty$$

a) Verify that the preceding is a joint density function.

b) Find $Cov(X,Y)$.

- To show that $f(x,y)$ is a joint density function we need to show it is nonnegative, which is immediate and that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx = 1$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dy\,dx = \int_{0}^{\infty}\int_{0}^{\infty} \frac{1}{y}e^{-(y+x/y)}dy\,dx$$

$$= \int_{0}^{\infty} e^{-y}\int_{0}^{\infty} \frac{1}{y}e^{-x/y}dx\,dy$$

$$= \int_{0}^{\infty} e^{-y}dy = 1$$

Covariance and Variance Example 1

- To obtain $Cov(X,Y)$, note that the density function of Y is

$$f_Y(y) = e^{-y}\int_0^\infty \frac{1}{y}e^{-x/y}dx = e^{-y}$$

- Thus Y is an exponential random variable with parameter 1

$$E[Y] = 1$$

- Compute $E[X]$ and $E[XY]$ as follows:

$$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dy\,dx$$

$$= \int_0^\infty e^{-y}\int_0^\infty \frac{x}{y}e^{-x/y}dx\,dy$$

- Now, $\int_0^\infty \frac{x}{y}e^{-x/y}dx$ is the expected value of an exponential random variable with parameter 1/y, and thus is equal to y. Consequently,

$$E[X] = \int_0^\infty ye^{-y}dy = 1$$

Covariance and Variance Example 1

- Also $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dy\,dx$

$$= \int_0^{\infty} ye^{-y}\int_0^{\infty} \frac{x}{y}e^{-x/y}dx\,dy$$

$$= \int_0^{\infty} y^2e^{-y}dy$$

- Integration by parts $(dv = e^{-y}dy, u = y^2)$ gives

$$E[XY] = \int_0^{\infty} y^2e^{-y}dy = -y^2e^{-y}\Big|_0^{\infty} + \int_0^{\infty} 2ye^{-y}dy = 2E[Y] = 2$$

- Consequently,

$$Cov(X,Y) = E[XY] - E[X]E[Y] = 1$$

Covariance and Variance Example 2 (Quiz 1)

Problem 2 - Suppose *n* fair dice are independently rolled. Let:

$$X_k = \begin{cases} 1 & \text{if 1 shows on the } k^{th} \text{ die} \\ 0 & \text{else} \end{cases} \qquad Y_k = \begin{cases} 1 & \text{if 2 shows on the } k^{th} \text{ die} \\ 0 & \text{else} \end{cases}$$

Let $X = \sum_{k=1}^{n} X_k$, which is the number of one's showing, and $Y = \sum_{k=1}^{n} Y_k$, which is the number of two's showing. Note that if a histogram is made recording the number of occurrences of each of the six numbers, then *X* and *Y* are the heights of the first two entries in the histogram.

a) Find *E[X]* and *Var(X)*.
b) Find *Cov(X_i, Y_j)* if $1 \le i \le n$ and $1 \le j \le n$ (**Hint:** Does it make a difference if i = j?)
c) Find *Cov(X, Y)*.
d) Find the correlation coefficient $\rho_{X,Y}$. Are *X* and *Y* positively correlated, uncorrelated, or negatively correlated?

Hint: X_k is a Bernoulli random variable with $p = \frac{1}{6}, E[X_k] = \frac{1}{6}$, and $Var(X_k) = \frac{5}{36}$.

Remember that:

$$Cov(X,Y) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, Y_j)$$

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Covariance and Variance Example 2 (Cont'd)

a) X_k is bernouli random variable with $p = \frac{1}{6}$ (Probability of one showing)

$$E[X_k] = 1(p) + 0(1-p) = \frac{1}{6}$$

$$Var(X_k) = E[X_k^2] - E[X_k]^2 = \frac{1}{6} - \frac{1}{36} = \frac{5}{36}$$

$$E[X_k^2] = (1)^2 p + (0)^2 (1-p) = \frac{1}{6}$$

b) $$E[X] = E\left[\sum_{k=1}^{n} X_k\right] = E[X_1] + E[X_2] + \cdots + E[X_k] = \frac{n}{6}$$

$$Var(X) = Var\left(\sum_{k=1}^{n} X_k\right) = Var(X_1) + Var(X_2) + \cdots + Var(X_k) = \frac{5n}{36}$$

c) if $i \neq j$ then X_i & Y_j are independent (different dices)

* $\Rightarrow$ $Cov(X_i, Y_j) = 0$

if $i = j$ then: * $Cov(X_i, Y_i) = E[X_i Y_i] - \overset{\frac{1}{6}}{E[X_i]}\, \overset{\frac{1}{6}}{E[Y_i]}$

joint PMF

$X_i Y_i \rightarrow$

$(0,0)$ $\frac{4}{6}$ $\Rightarrow$ $\boxed{E[X_i Y_i]} = \sum_{i=0,1} \sum_{j=0,1} i j P_{ij} = 0 \times 0 \left(\frac{4}{6}\right) + 0 \times 1 \left(\frac{1}{6}\right) + 1 \times 0 \left(\frac{1}{6}\right) = 0$

* $(0,1)$ $\frac{1}{6}$

$(1,0)$ $\frac{1}{6}$

not possible ~~(1,1)~~

$$\boxed{Cov(X_i, Y_i) = -\frac{1}{36}}$$

d) $$Cov(X,Y) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, Y_j) = \sum_{i=1}^{n} Cov(X_i, Y_i) + \overset{0}{\sum_{i \neq j} Cov(X_i, Y}$$

$$\boxed{Cov(X,Y) = -\frac{n}{36}}$$

Central Limit Theorem Example (Quiz 1)

- **Problem 1 -** From past experience a professor knows that the test score of each student taking final examination is a random variable with a mean of 75 and standard deviation of 8. How many students would have to take the examination to ensure, with probability at least .95, that the class mean would be at least 73?

- **Hint:** As discussed in the class, the central limit theorem applies to both the sum of a sequence of independent and identically distributed random variables, and their mean.

Central Limit Theorem Example (Cont'd)

$$* \left\{ \begin{array}{l} \bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n} \\ E[\bar{X}] = \frac{1}{n} E[X_1] + E[X_2] + \cdots + E[X_n] = E[X_1] = 75 \\ Var(\bar{X}) = Var\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right) = \dfrac{1}{n^2}\left[Var(X_1) + \cdots + Var(X_n)\right] = \dfrac{Var(X_1)}{n} = \dfrac{64}{n} \end{array} \right.$$

$$\text{CLT} \Rightarrow \bar{X} \sim \mathcal{N}\left(75, \frac{8}{\sqrt{n}}\right)$$

$$P(\bar{X} \geqslant 73) = P\left(\frac{\bar{X} - 75}{\frac{8}{\sqrt{n}}} \geqslant \frac{73 - 75}{\frac{8}{\sqrt{n}}}\right) \approx \underbrace{P\left(Z \geqslant \frac{-\sqrt{n}}{4}\right) \geqslant 0.95}_{*}$$

$$1 - \underbrace{\Phi\left(-\frac{\sqrt{n}}{4}\right)}_{* \; 1 - \Phi\left(\frac{\sqrt{n}}{4}\right)} \geqslant 0.95 \Rightarrow \Phi\left(\frac{\sqrt{n}}{4}\right) \geqslant 0.95$$

$$\frac{\sqrt{n}}{4} \geqslant \underbrace{1.65}_{\text{Needs table}} \Rightarrow \sqrt{n} \geqslant 6.6$$

$$n \geqslant (6.6)^2 = 43.56$$

$$n \text{ is an integer} \Rightarrow \boxed{n = 44}$$

We need at least 44 students taking the exam.