# STATISTICS 2

NOTES BY
GAGNEET
KAUR

September 15, 2021

# WEEK 2

## Independence of Two Random Variables

\# <u>Definition</u> : Let X and Y be two random variables defined in a probability space with ranges $T_X$ and $T_Y$, respectively. X and Y are said to be independent if any event defined using X alone is independent of any event defined using Y alone.

Equivalently, if the joint PMF of X and Y is $f_{XY}$, X & Y are independent if

$$f_{XY}(t_1, t_2) = f_X(t_1) \cdot f_Y(t_2) \quad ; \quad t_1 \in T_X,\ t_2 \in T_Y$$

\# <u>General Case Vs Independent Case</u>

General : $f_{XY}(t_1, t_2) = f_X(t_1) \cdot f_{Y|X=t_1}(t_2)$

Independent : $f_{XY}(t_1, t_2) = f_X(t_1) \cdot f_Y(t_2)$

& $\underbrace{f_{Y|X=t_1}(t_2)}_{\text{conditional PMF}} = \underbrace{f_Y(t_2)}_{\text{marginal PMF}}$

- If X and Y are independent
  - → Joint PMF equals product of marginal PMFs
  - → Conditional PMF equals marginal PMF

# INDEPENDENT VS DEPENDENT R.V.

- To show X and Y are independent, verify

$$f_{XY}(t_1, t_2) = f_X(t_1) \cdot f_Y(t_2)$$

for <u>all</u> $t_1 \in T_X$, $t_2 \in T_Y$.

- To show X and Y are dependent, verify

$$f_{XY}(t_1, t_2) \neq f_X(t_1) \cdot f_Y(t_2)$$

for <u>some</u> $t_1 \in T_X$ and $t_2 \in T_Y$.

- <u>Special case</u> : $f_{XY}(t_1, t_2) = 0$ when $f_X(t_1) \neq 0$ and $f_Y(t_2) \neq 0$.

# Independence of Multiple Random Variables

<u>Definition</u>: Let $X_1, \ldots, X_n$ be random variables defined in a probability space with range of $X_i$ denoted $T_{X_i}$. $X_1, \ldots, X_n$ are said to be independent if events defined using different $X_i$ are mutually independent.

Equivalently, $X_1, \ldots, X_n$ are independent iff

$$\boxed{f_{X_1 \ldots X_n}(t_1, \ldots, t_n) = f_{X_1}(t_1) \cdot f_{X_2}(t_2) \ldots f_{X_n}(t_n)}$$

for all $t_i \in T_{X_i}$.

# If X and Y are independent
- → Joint PMF equals product of marginal PMFs
- → Conditional PMF equals unconditioned PMF.

# All subsets of independent random variables are independent.

---

<u>Example</u>: Random 3-digit number 000 to 999

X = first digit from left
Y = no. modulo 2
Z = first digit from right

- $X \sim$ Uniform $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $Y \sim$ uniform $\{0, 1\}$
- $Z \sim$ Uniform $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

→ X, Z
- $f_{XZ}(t_1, t_3) = 1/100 = f_X(t_1) \cdot f_Z(t_3)$
- Independent

→ X, Y
- $f_{XY}(t_1, t_2) = 1/20 = f_X(t_1) \cdot f_Y(t_2)$
- Independent

→ Y, Z
- $f_{YZ}(t_2, t_3) = 0 \neq f_Y(1) f_Z(2)$
- Not independent

---

# INDEPENDENT & IDENTICALLY DIST-RIBUTED (i.i.d)

__Definition__: Random variables $X_1, \ldots, X_n$ are said to be independent and identically distributed (i.i.d.), if

(1) they are independent
(2) the marginal PMFs $f_{X_i}$ are identical.

\# Repeated trials of an experiment creates i.i.d. sequence of random variables.
→ Toss a coin multiple times

- Throw a die multiple times

# Notation : $X_1, X_2, \ldots, X_n$ are i.i.d. with distribution $X$ means that $f_{X_i}$ is the same as $f_X$.

$$\boxed{X_1, X_2, \ldots, X_n \sim \text{i.i.d. } X}$$

---

# Problem : i.i.d. Geometric

Let $X_1, \ldots, X_n$ be i.i.d. with a Geometric($p$) distribution. What is the probability that all of these random variables are larger than some positive integer $j$?

$X \sim \text{Geometric}(p) : X \in \{1, 2, 3, \ldots\}$

$$Pr(X=k) = (1-p)^{k-1} p$$

We have to find,

$$P(X_1 > j, X_2 > j, \ldots, X_n > j) = P(X_1 > j) \cdot P(X_2 > j) \ldots P(X_n > j)$$
$$= [P(X > j)]^n$$

Now, $P(X > j) = \sum_{k=j+1}^{\infty} (1-p)^{k-1} p = \frac{(1-p)^j p}{1-(1-p)} = (1-p)^j$

$\therefore \; P(X_1 > j, X_2 > j, \ldots, X_n > j) = [P(X > j)]^n = (1-p)^{jn}$

---

# Problem : i.i.d samples

Let $X \sim \{\overset{1/2}{0}, \overset{1/4}{1}, \overset{1/8}{2}, \overset{1/16}{3}, \overset{1/16}{4}\}$, and let $X_1, \dots, X_n$ be i.i.d samples with distribution $X$.

(1) What is the probability that 4 is missing in the samples?

(2) What is the probability that 4 appears exactly once in the samples?

(3) What is the probability that 3 and 4 appears at least once in the samples? (doubt)

Solution: (1) $P(X_1 \neq 4, X_2 \neq 4, \dots, X_n \neq 4) = [P(X \neq 4)]^n = \left(\frac{15}{16}\right)^n$

(2) $$P(\text{4 exactly appears once}) = P(X_1 = 4, X_2 \neq 4, \dots, X_n \neq 4) + P(X_1 \neq 4, X_2 = 4, \dots, X_n \neq 4) + \dots + P(X_1 \neq 4, X_2 \neq 4, \dots, X_{n-1} \neq 4, X_n = 4)$$

$$= n\left(\frac{1}{16}\right)\left(\frac{15}{16}\right)^{n-1}$$

(3) ~~If~~ Let A be the event → 3 appears at least once
& B be the event → 4 appears at least once

We need to find, $P(A \cap B)$

Now, $A^c$ = 3 never appears $\Rightarrow P(A^c) = (15/16)^n$
$B^c$ = 4 never appears $\Rightarrow P(B^c) = (15/16)^n$

$$P(A^c \cup B^c) = P(A^c) + P(B^c) - P(A^c \cap B^c)$$
$$= \left(\frac{15}{16}\right)^n + \left(\frac{15}{16}\right)^n - \left(\frac{14}{16}\right)^n$$

$$\Rightarrow P(A^c \cup B^c) = 2\left(\frac{15}{16}\right)^n - \left(\frac{14}{16}\right)^n$$

Thus, $P(A \cap B) = 1 - P(A^c \cup B^c)$

$$\Rightarrow P(A \cap B) = 1 - 2\left(\frac{15}{16}\right)^n + \left(\frac{14}{16}\right)^n$$

(probability that 3 & 4 appear at least once)

---

# MEMORYLESS PROPERTY OF GEOMETRIC

Let $X \sim$ Geometric $(p)$. Find the following:

(1) $P(X > n)$

(2) $P(X > m+n \mid X > m)$

$$\rightarrow P(X > n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p = (1-p)^n$$

$$\rightarrow P(X > m+n \mid X > m) = \frac{P(X > m+n \cap X > m)}{P(X > m)} = \frac{P(X > m+n)}{P(X > m)}$$

$$= \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n$$

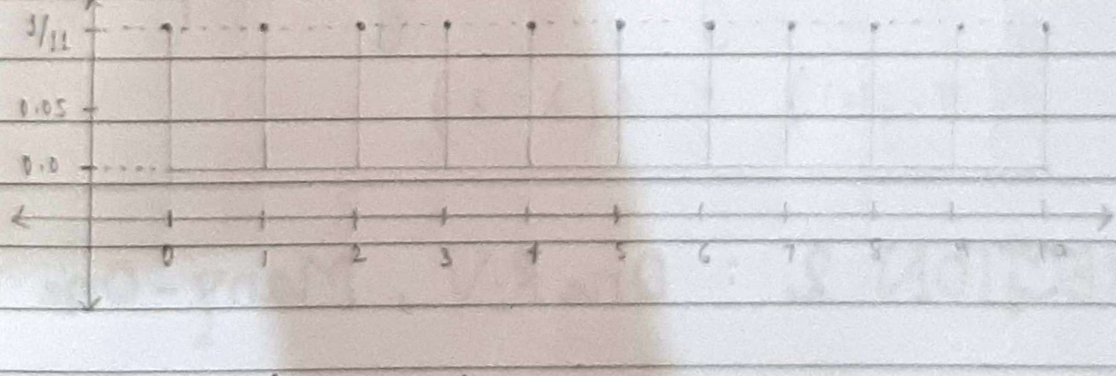
We conclude that,

$$\boxed{P(X > m+n \mid X > m) = P(X > n)}$$

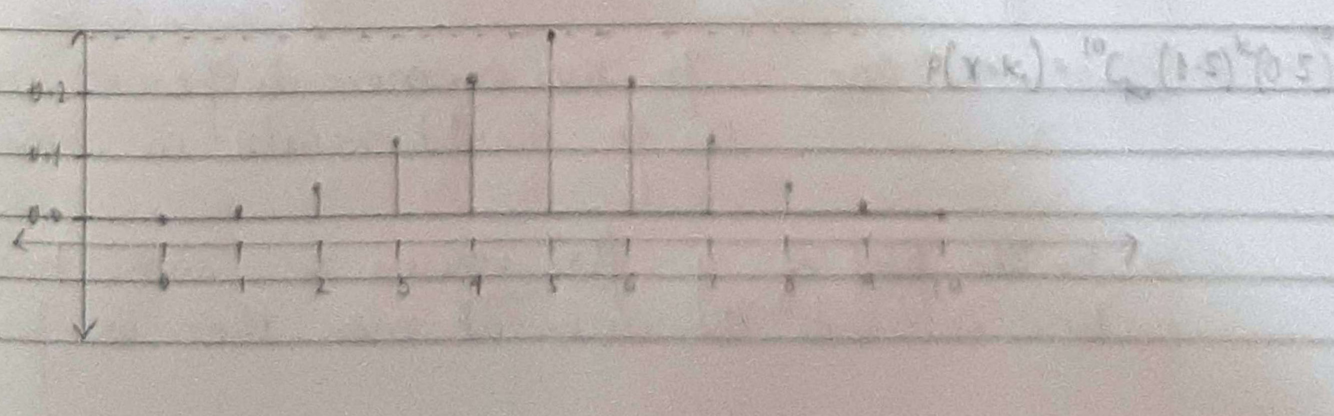
# Visualizing Functions of One Random Variable

## SECTION 1 : One R.V. , One-to-One $F^n s$

(1) Uniform $\{0, 1, 2, \ldots, 10\}$




(2) Binomial $(10, 0.5)$



## One-to-One Functions

- One-to-one : For different input, the $f^n$ results in different outputs.

- In the plot of $y = f(x)$ versus $x$, every horizontal line intersects the curve only at one point.

- Monotonic $f^n$ are one-to-one.

- Example : $x-5$, $2^x$.

- If $f$ is one-to-one,

- Easy to find PMF of $f(x)$ using the table method.

- Easy to visualize the plot of the PMF.

$$P(Y = f(x)) = P(X = x)$$

---

# SECTION 2 : One R.V, Many-One $F^n$s

## Many-to-One Functions

- Many-to-One : For two or more different inputs, the $f^n$ should result in same output.

- In the plot of $y = f(x)$ versus $x$, some horizontal line intersects the curve at more than one point.

- Non-monotonic $f^n$ are many-to-one.

- Example : $y = x^2$ , $y = x(1-x)$ , $y = xe^{-x}$

* <u>If $f$ is many-to-one :</u>

- The table method will generally work.

- Let $y_0$ be value taken by $f$ at the points $x_1, x_2, \ldots, x_m$ and nowhere else.
  That is, $y_0 = f(x_1) = \ldots = f(x_m)$

$$P(Y = y_0) = P(X = x_1) + \ldots + P(X = x_m)$$

Date: September 17, 2021

classmate
Date
Page

# EXAMPLES ON FUNCTIONS OF ONE RANDOM VARIABLE

(1) Let $X \sim \text{Uniform}\{-500, -499, \ldots, 500\}$. Let $f(x) = \max(x, 5)$. Find the distribution of $Y = f(X)$.

Sol.

$$f(x) = \max(x, 5)$$

$$\Rightarrow f(x) = \begin{cases} x, & x \geq 5 \\ 5, & x < 5 \end{cases}$$

We have, $X \sim \text{Uniform}\{-500, -499, \ldots, 500\}$

$Y = f(X) \Rightarrow Y \in \{5, 6, 7, \ldots, 500\}$

$P(5) = \frac{506}{1001}$ and $P(6) = P(7) = \ldots = P(500) = \frac{1}{1001}$.

# Introduction to functions of two Random Variables

## # Small Examples and Table Method

(1) Sum

$X, Y \sim$ i.i.d. Uniform $\{0, 1\}$ , $Z = X + Y$

| X | Y | $f_{XY}(x, y)$ | Z |
|---|---|---|---|
| 0 | 0 | 1/4 | 0 |
| 0 | 1 | 1/4 | 1 |
| 1 | 0 | 1/4 | 1 |
| 1 | 1 | 1/4 | 2 |

$Z \in \{0, 1, 2\}$

$P(Z=0) = 1/4$ ; $P(Z=1) = 1/2$ ; $P(Z=2) = 1/4$

(2) Max

$Z = \max(X, Y)$

$P(Z=0) = 1/2$
$P(Z=1) = 11/32$
$P(Z=2) = 5/32$

| X | Y | $f_{XY}(x, y)$ | Z |
|---|---|---|---|
| 0 | 0 | 1/2 | 0 |
| 0 | 1 | 1/4 | 1 |
| 0 | 2 | 1/8 | 2 |
| 1 | 0 | 1/16 | 1 |
| 1 | 1 | 1/32 | 1 |
| 1 | 2 | 1/32 | 2 |

# Moderate Size Examples : Too Cumbersome

- Pair of fair dice are thrown. What is the distribution of the sum or max or min?

- Table method
  → 36 possible values
  → Prone to errors

- What if size becomes 100?

---

# VISUALISING FUNCTIONS OF TWO VARIABLES

- $g(x,y)$ : function
  → 3D plot is one option, but often not very useful

- Contours : values of $(x,y)$ that result in $g(x,y) = c$
  → Make a plot of those $(x,y)$ for different $c$

- Regions : values of $(x,y)$ that result in $g(x,y) \leq c$
  → Make a plot of those $(x,y)$ for different $c$

---

# FUNCTION OF TWO R.V.s

$X, Y \sim f_{XY}$, $X \in \mathcal{X}$, $Y \in \mathcal{Y}$

- Let $Z = g(X,Y)$ be a function of $X$ and $Y$.

- What is the PMF of Z ?

Step 1 : Find range of Z

→ All possible values for $g(x,y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$

Step 2 : Add over the contours

→ Suppose $z$ is a possible value taken by Z

$$P(Z=z) = \sum_{(x,y):\, g(x,y)=z} f_{XY}(x,y)$$

---

EXAMPLE : Pair of Dice

1. Sum

$X, Y \sim$ iid Unif $\{1,2,3,4,5,6\}$, $Z = X+Y$

Step 1 : Find the range of Z

→ $Z \in \{2,3,4,5,6,7,8,9,10,11,12\}$

Step 2 : Add over the contours

| z | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P(Z=z) | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |

2. Max

$X, Y \sim$ iid Unif $\{1,2,3,4,5,6\}$ , $Z = \max(X,Y)$

Step 1 : Find range of $Z$

$\rightarrow Z \in \{1, 2, 3, 4, 5, 6\}$

Step 2 : Add over the contours

| $z$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $P(Z=z)$ | $1/36$ | $3/36$ | $5/36$ | $7/36$ | $9/36$ | $11/36$ |

# GENERALISATION

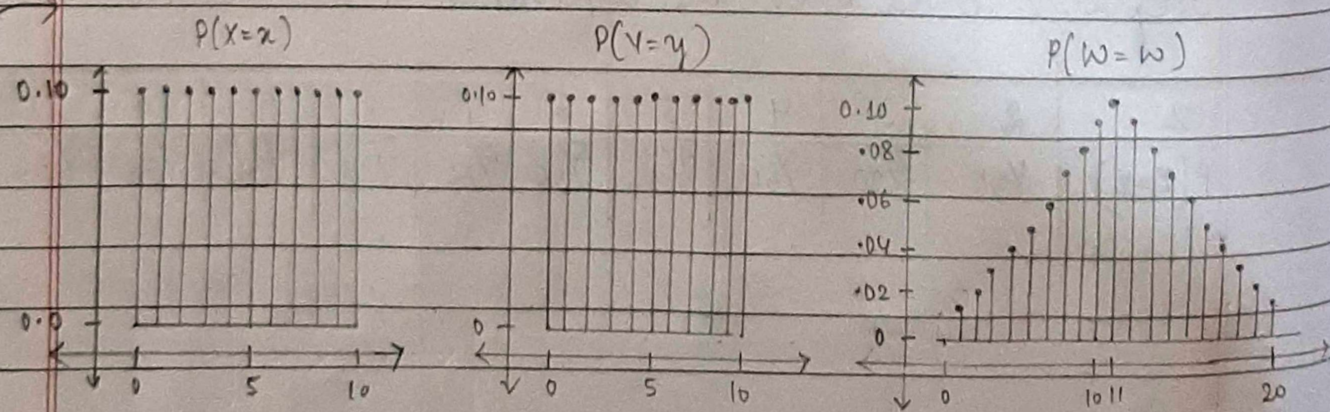
(1) iid Uniform $\{1, 2, \ldots, n\}$ : SUM

$X, Y \sim$ iid Unif $\{1, 2, 3, \ldots n\}$ , $W = X + Y$

$W \in \{2, 3, \ldots, 2n\}$

$$\therefore \; P(W=w) = \begin{cases} w-1/n^2 , & 2 \le w \le n+1 \\ \dfrac{2n-w+1}{n^2} , & n+2 \le w \le 2n \end{cases}$$

$\boxed{n=10}$



(2) iid Uniform $\{1, 2, \ldots, n\}$ : Max

$X, Y \sim$ iid Unif $\{1, 2, \ldots, n\}$, $Z = \max(X, Y)$

$Z = \{1, 2, \ldots n\}$

$\therefore P(Z = z) = \dfrac{2z - 1}{n^2}$

September 19, 2021

# Functions of Random Variables

## PMF of $g(X_1, \ldots, X_n)$

Suppose $X_1, \ldots, X_n$ have joint PMF $f_{X_1 \ldots X_n}$ with $T_{X_i}$ denoting the range of $X_i$.
Let $g : T_{X_1} \times \ldots \times T_{X_n} \to \mathbb{R}$ be a function with range $T_g$. The PMF of $X = g(X_1, \ldots, X_n)$ is given by

$$f_X(t) = P(g(X_1, \ldots, X_n) = t) = \sum_{(t_1, \ldots t_n) : g(t_1 \ldots t_n) = t} f_{X_1 \ldots X_n}(t_1, \ldots, t_n)$$

→ Proof : write the event & use definition of joint PMF
→ Directly useful for small problems
→ Can be extended for joint PMF of two functions g and h.

---

<u>Example</u> : Binomial from Bernoulli Trials

Let $X_1, \ldots, X_n$ be the results of $n$ i.i.d. Bernoulli$(p)$ trials. The sum of the $n$ random variables $X_1 + X_2 + \ldots + X_n$ is Binomial$(n, p)$.

i.e, Sum of $n$ independent Bernoulli $(p)$ = Binomial$(n, p)$

**Example**: sum of two uniforms

Let $X \sim$ Uniform $\{0,1,2,3\}$ and $Y \sim$ Uniform $\{0,1,2,3\}$ be independent. Find the PMF of $Z = X+Y$.

$$Z \in \{0, 1, 2, 3, 4, 5, 6\}$$

$$P(Z=z) \quad \tfrac{1}{16} \quad \tfrac{2}{16} \quad \tfrac{3}{16} \quad \tfrac{4}{16} \quad \tfrac{3}{16} \quad \tfrac{2}{16} \quad \tfrac{1}{16}$$

PMF Table :

| $z$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $P(Z=z)$ | 1/16 | 2/16 | 3/16 | 4/16 | 3/16 | 2/16 | 1/16 |

## SUMS OF TWO R.V. TAKING INT. VALUES

Suppose $X$ and $Y$ take integer values and let their joint PMF be $f_{XY}$. Let $Z = X+Y$.

Let $z$ be some integer.

$$
\begin{aligned}
P(Z=z) &= P(X+Y=z) \\
&= \sum_{x=-\infty}^{\infty} P(X=x, Y=z-x) \\
&= \sum_{x=-\infty}^{\infty} f_{XY}(x, z-x) \\
&= \sum_{y=-\infty}^{\infty} f_{XY}(z-y, y)
\end{aligned}
$$

Convolution : If $X$ & $Y$ are independent,

$$f_{X+Y}(z) = \sum_{x=-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)$$

# SUM OF TWO INDEPENDENT POISSONS

Let $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{poisson}(\lambda_2)$ be independent.

1. Find PMF of $Z = X+Y$
2. Find conditional distribution of $X|Z$.

(1) $Z \in \{0, 1, 2, \ldots\}$

$$f_Z(z) = \sum_{x=0}^{z} f_X(x) \cdot f_Y(z-x)$$

$$= \sum_{x=0}^{z} \frac{e^{-\lambda_1} \lambda_1^x}{x!} \cdot \frac{e^{-\lambda_2} \lambda_2^{z-x}}{(z-x)!}$$

$$\Rightarrow f_Z(z) = \frac{e^{-\lambda_1} \cdot e^{-\lambda_2}}{z!} \sum_{x=0}^{z} \frac{z!}{x!\,(z-x)!} \lambda_1^x \cdot \lambda_2^{z-x}$$

$$\Rightarrow f_Z(z) = \frac{e^{-(\lambda_1+\lambda_2)} \cdot (\lambda_1 + \lambda_2)^z}{z!}$$

$\therefore Z \sim \text{poisson}(\lambda_1 + \lambda_2)$

(2) $P(X=k \mid Z=n) = \dfrac{P(X=k, Z=n)}{P(Z=n)}$

$$= \frac{P(X=k) \cdot P(Z=n \mid X=k)}{P(Z=n)}$$

$$= \frac{P(X=k) \cdot P(Y=n-k)}{P(Z=n)}$$

$$= \frac{(e^{-\lambda} \cdot \lambda_1^k)/k! \times (e^{-\lambda_2} \cdot \lambda_2^{n-k})/(n-k)!}{\left(\frac{e^{-(\lambda_1+\lambda_2)} \cdot (\lambda_1+\lambda_2)^n}{n!}\right)}$$

$$\Rightarrow P(X=k \mid Z=n) = \frac{n!}{k!(n-k)!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

$\therefore$ $X \mid Z \sim$ Binomial $(n, \lambda_1/\lambda_1+\lambda_2)$
$Y \mid Z \sim$ Binomial $(n, \lambda_2/\lambda_1+\lambda_2)$

---

# FUNCTIONS AND INDEPENDENCE

- If $X$ and $Y$ are independent, $g(X)$ and $h(Y)$ are independent for any two $f^n$ $g$ & $h$.

- If $X_1, X_2, X_3, X_4$ are mutually independent,
  - $\rightarrow$ $g(X_1, X_2)$ is independent of $h(X_3, X_4)$
  - $\rightarrow$ $g(X_1, X_2, X_3)$ is independent of $h(X_4)$

- functions of non-overlapping sets of independent random variables are also independent.

---

# Min. & Max. of Two R.V.

## MINIMUM OF 2 R.Vs

$X, Y \sim f_{XY}$

$Z = \min(X, Y)$ : function of $X$ and $Y$

Example : Throw a dice twice
$Z$ = min. of 2 no. seen

$f_Z(z) = P[\min(X, Y) = z]$

$= P((X = z \text{ and } Y = z) \text{ or } (X = z \text{ and } Y > z) \text{ or } (X > z \text{ and } Y = z))$

$\Rightarrow f_Z(z) = f_{XY}(z, z) + \sum_{t_2 > z} f_{XY}(z, t_2) + \sum_{t_1 > z} f_{XY}(t_1, z)$

---

## Independent Case : CDF of maximum

CDF of a R.V : Cummulative Distribution function of a random variable $X$ is a function $F_X : \mathbb{R} \to [0, 1]$ defined as

$$\boxed{F_X(x) = P(X \le x)}$$

Suppose $X$ and $Y$ are independent and $Z = \max(X, Y)$

$$\begin{aligned} F_Z(z) &= P(\max(X,Y) \le z) \\ &= P((X \le z) \text{ and } (Y \le z)) \\ &= P(X \le z) \cdot P(Y \le z) \\ &= F_X(z) \cdot F_Y(z) \end{aligned}$$

Independent : CDF of maximum product of CDFs.

---

## Problem : Min & Max of i.i.d. sequences

Let $X_1, \ldots, X_n \sim$ i.i.d $X$. Find the distribution of the following :

1. $\min(X_1, \ldots, X_n)$
2. $\max(X_1, \ldots, X_n)$

Solution: (1) $P(\min(X_1, \ldots, X_n) \ge z) = P(X_1 \ge z, X_2 \ge z, \ldots, X_n \ge z)$
$= [P(X \ge z)]^n$

(2) $P(\max(X_1, \ldots, X_n) \le z) = P(X_1 \le z, X_2 \le z, \ldots, X_n \le z)$
$= [P(X \le z)]^n = [F_X(z)]^n$

---

## Problem : Min. of 2 independent Geometrics

Let $X \sim$ Geometric$(p)$ and $Y \sim$ Geometric$(p)$ be independent. Find the distribution of $\min(X, Y)$.

Solution: $P(\min(X,Y) \geq k) = P(X \geq k) \cdot P(Y \geq k)$
$= (1-p)^{k-1} \cdot (1-p)^{k-1}$
$= (1-p)^{2(k-1)}$

$P(\min(X,Y) \geq k+1) = P(X \geq k+1) \cdot P(Y \geq k+1)$
$= (1-p)^{k} \cdot (1-p)^{k}$
$= (1-p)^{2k}$

$\therefore P(\min(X,Y) = k) = P(\min(X,Y) \geq k) - P(\min(X,Y) \geq k+1)$
$= (1-p)^{2(k-1)} - (1-p)^{2k}$

Let $(1-p)^2 = q$ (assume)

$\rightarrow P(\min(X,Y) = k) = q^{k-1} - q^{k} = q^{k-1}(1-q)$

$\therefore \min(X,Y) \sim \text{Geometric}(1-q)$

If $X \sim \text{Geometric}(p_1)$
$Y \sim \text{Geometric}(p_2)$ independently,

$\therefore \min(X,Y) \sim \text{Geometric}(1-(1-p_1)(1-p_2))$

or $\min(X,Y) \sim \text{Geometric}(p_1 + p_2 - p_1 p_2)$