BSCMA1004

STATISTICS II NOTES

WEEK 2 NOTES

IITM B.S Degree

PREPARED BY

Vehaan Handa, IIT Madras

Statistics II → Week II Notes

Independence of Two Random Variables

Let X and Y be 2 random variables defined in a probability space with range T_X, T_Y. X and Y are said to be independent if any event defined using X alone is independent of any event defined using Y alone. Or :→

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

$t_1 \in T_X,\ t_2 \in T_Y$

General: $f_{XY}(t_1,t_2) = f_X(t_1) f_{Y|X=t_1}(t_2)$
Independent: $f_{Y|X=t_1}(t_2) = f_Y(t_2)$.

If X and Y are independent:→
- Joint PMF = product of marginal PMFs
- Conditional PMF = Marginal PMF.

Eg.

$t_2 \backslash t_1$	0	1	f_Y
0	x	$1/2 - x$	$1/2$
1	$1/2 - x$	x	$1/2$
f_X	$1/2$	$1/2$	

$x = 1/4$, indep
$x \neq 1/4$, not indep.

$t_2 \backslash t_1$	0	1	2	f_Y
0	$1/6$	$1/12$	$1/12$	$\boxed{1/3}$
1	$1/4$	$1/8$	$1/8$	$\boxed{1/2}$
2	$1/12$	$1/24$	$1/24$	$\boxed{1/6}$
f_X	$\boxed{1/2}$	$\boxed{1/4}$	$\boxed{1/4}$	

independent

For independence, $f_{XY}(t_1, t_2) = f_X(t_1) f_Y(t_2) \quad \forall\, t_1, t_2 \in T_X, T_Y$

For dependence, $f_{XY}(t_1, t_2) \neq f_X(t_1) f_Y(t_2)$ for some $t_1, t_2 \in T_X, T_Y$

Special case: $f_{XY}(t_1, t_2) = 0$ when $f_X(t_1) \neq 0$, $f_Y(t_2) \neq 0$.

Eg. X = no. of runs in over
Q Y = no. of wickets in over. Are X & Y independent?

It would seem like ... no, not really.

★ Independence of Multiple Random Variables

$$f_{X_1 \ldots X_n}(t_1, \ldots, t_n) = f_{X_1}(t_1) f_{X_2}(t_2) \ldots f_{X_n}(t_n). \quad \forall\, t_i \in T_{X_i}$$

All subsets of independent RV's are independent.

Eg. Even Parity

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1, t_2, t_3)$
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	0	1/4

No. of 1's in (X_1, X_2, X_3) is even, hence the name.
$X_i \sim$ Uniform $\{0, 1\}$
All pairs independent.
$f_{X_1X_2X_3}(0, 0, 1) = 0 \neq f_{X_1}(0) f_{X_2}(0) f_{X_3}(1)$

X_1, X_2, X_3 : Dependent

DATE

★ Independent and Identically Distributed (iid)

Random variables $X_1, X_2, \ldots, X_n$ are said to be independent and identically distributed (i.i.d) if:→
① they are independent
② 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.

$$X_1, X_2, \ldots, X_n \sim \text{iid } X$$

Eg. Let $X_1, \ldots, X_n$ be i.i.d with a Geometric (p) distribution. What is the probability that all these rv's are larger than some +ve integer j?

$p \quad (1-p)p \quad (1-p)^2 p.$

$X \in \{1, 2, 3, \ldots\}$ $\qquad P(X=k) = (1-p)^{k-1} p.$

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

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

Ans. $(1-p)^{jn}$

classmate

DATE

Q. Let $X \sim \{0, 1, 2, 3, 4\}$ with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}$, let $X_1 \ldots X_n$ be iid samples, dist. X.

1. Pr(4 missing in the samples)?
2. Pr(4 appears exactly once in the samples)?
3. Pr(3 & 4 appear at least once in the samples)?

1. $P(X_1 \neq 4, X_2 \neq 4, \ldots, X_n \neq 4) = (P(X \neq 4))^n = \left(\frac{15}{16}\right)^n$

2. $P(4 \text{ app. exactly once}) = P((X_1 = 4), (X_2 \neq 4), (\ldots X_n \neq 4))$
$+ P(X_1 \neq 4, X_2 = 4, \ldots X_n \neq 4)$
$+ \ldots + \ldots + P(X_1 \neq 4 \ldots, X_n = 4)$

$= n \cdot P(X = 4)\left((P(X \neq 4)\right)^{n-1}$
$= \boxed{n \cdot \left(\frac{1}{16}\right)\left(\frac{15}{16}\right)^{n-1}}$

3. $P(\underset{A}{3 \text{ atleast once}} \cap \underset{B}{4 \text{ atleast once}})$ $\quad A \cap B = (A^c \cup B^c)^c$

$P(A^c) = \left(\frac{15}{16}\right)^n, \; P(B^c) = \left(\frac{15}{16}\right)^n \quad P(A^c \cap B^c) = \left(1 - \frac{1}{16} - \frac{1}{16}\right)^n = \left(\frac{14}{16}\right)^n$

$\boxed{P(A \cap B) = 1 - \left(2\left(\frac{15}{16}\right)^n - \left(\frac{14}{16}\right)^n\right)}$ $\quad P(A^c \cup B^c) = 2(15/16)^n - (14/16)^n$

★ Memoryless Property of Geometric (p)

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

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

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

If 1000 tosses have not got head, prob(waiting for +100) = starting at Toss 0 and Pr(first head after 100 tosses)

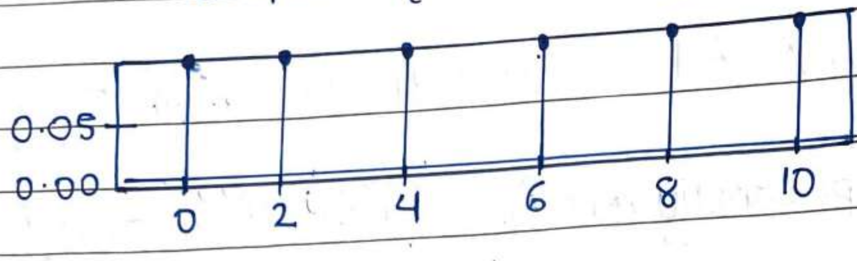
$\boxed{P(X > m+n \mid X > m) = P(X > n)}$

classmate PAGE

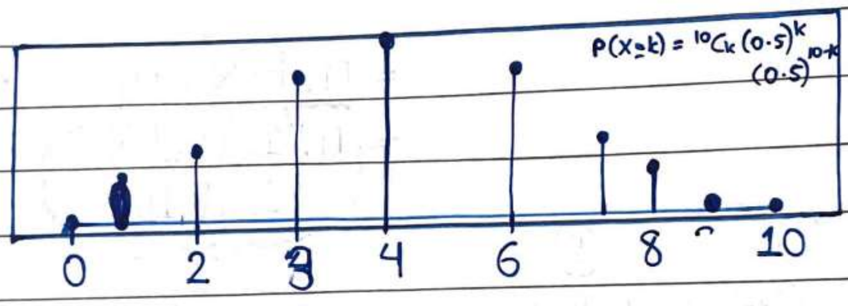
★ Visualising Functions of Random Variables

① 1 RV, 1 to 1 functions

Uniform $\{0, 1, \ldots, 10\}$



Binomial$(10, 0.5)$



x	$P(X=x)$		$y = x-5$
0	$1/11$	← Probability of $Y=-5$	-5
1	$1/11$		-4
2	$1/11$		-3
3	$1/11$		-2
4	$1/11$		-1
5	$1/11$		0
6	$1/11$	← Probability of $Y=1$	1
7	$1/11$		2
8	$1/11$		3
9	$1/11$		4
10	$1/11$		5
	~~1/11~~		
	~~1/11~~		

Uniform $\{0, 1, \ldots, 10\}$

Table Method

So the PMF remains same, only the values along the axis keep changing.

Eg. PMF of $Y = 2^x$

Binomial (10, 0.5)

The plot will appear bunched for smaller and spread out for larger values.

x	P(X=x)	$y=2^x$
0	0.00097656	1
1	0.00976563	2
2	0.04394531	4
3	0.1171875	8
4	0.20507813	16
5	0.24609375	32
6	0.20507813	64
7	0.1171875	128
8	0.04394531	256
9	0.00976563	512
10	0.00097656	1024

Monotonic fns, one to one. eg. $x-5$, 2^x.
Easy to find and visualise PMF.

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

② Many to 1 functions

x	P(X=x)	$y=(x-5)^2$
0	1/11	25
1	1/11	16
2	1/11	9
3	1/11	4
4	1/11	1
5	1/11	0
6	1/11	1
7	1/11	4
8	1/11	9
9	1/11	16
10	1/11	25

y	P(Y=y)
0	1/11
1	2/11
4	2/11
9	2/11
16	2/11
25	2/11

DATE

If f is many to 1:→
- Table generally works.

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

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

Q. Let $X \sim \text{Uniform}\{-5, -4, \ldots, 5\}$ Let:→

$f(x) = \begin{cases} x, & x > 0 \\ 0 & x \leq 0 \end{cases}$ Distribution of $Y = f(X)$.

x	$P(X=x)$	$y=f(x)$
-5	$1/11$	0
-4	$1/11$	0
-3	$1/11$	0
-2	$1/11$	0
-1	$1/11$	0
0	$1/11$	0
1	$1/11$	1
2	$1/11$	2
3	$1/11$	3
4	$1/11$	4
5	$1/11$	5

$Y \in \{0, 1, 2, 3, 4, 5\}$.

y	$P(Y=y)$
0	$6/11$
1	$1/11$
2	$1/11$
3	$1/11$
4	$1/11$
5	$1/11$

★ Functions of 2 RV's

$X, Y \sim$ iid Unif $\{0,1\}$, $Z = X+Y$

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

$Z = \{0,1,2\}$
$P(Z=0) = 1/4$
$P(Z=1) = 1/2$
$P(Z=2) = 1/4$

Eg. Pair of fair dice thrown. Distribution of sum/max or min?

Table method is cumbersome.

How to visualise?

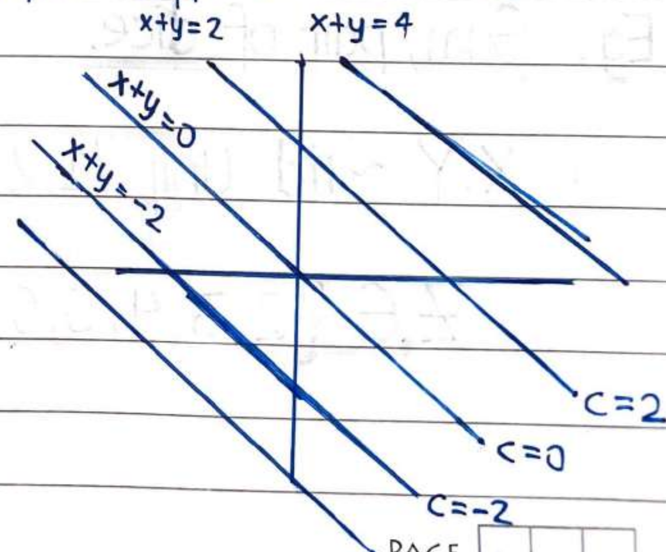
① One option is a 3D plot, however not very useful.

② Another option is contours.

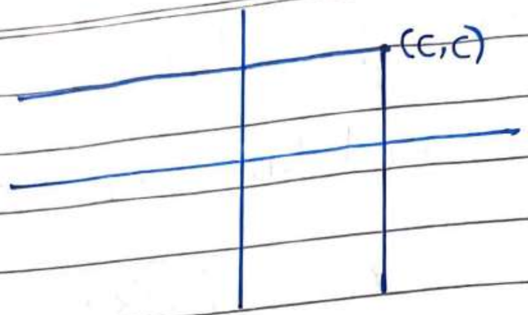
Contours: Values of (x,y) resulting in $g(x,y) = c$
- Make a plot of those (x,y) for diff. c.

[Or the values of (x,y) resulting in $g(x,y) \leq c$] → REGIONS
- Make a plot of (x,y) for diff. c.

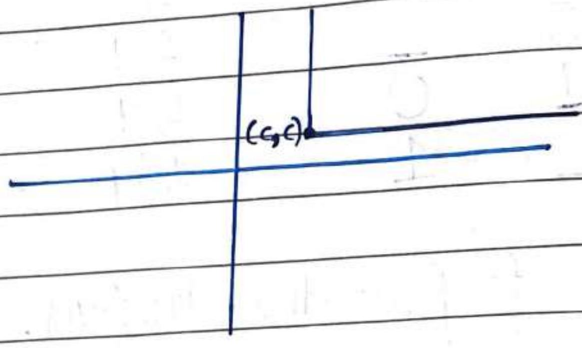
Eg. Sum fn $(x+y)$.



Eg. $\max(x,y) = c$



$\min(x,y) = c$



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

Let $Z = g(X,Y)$ is fn of **X** and **Y**. PMF of Z?

① Find range of Z

② Add over the contours :→

- Suppose z is a possible value taken by Z.

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

Eg. Sum, pair of dice.

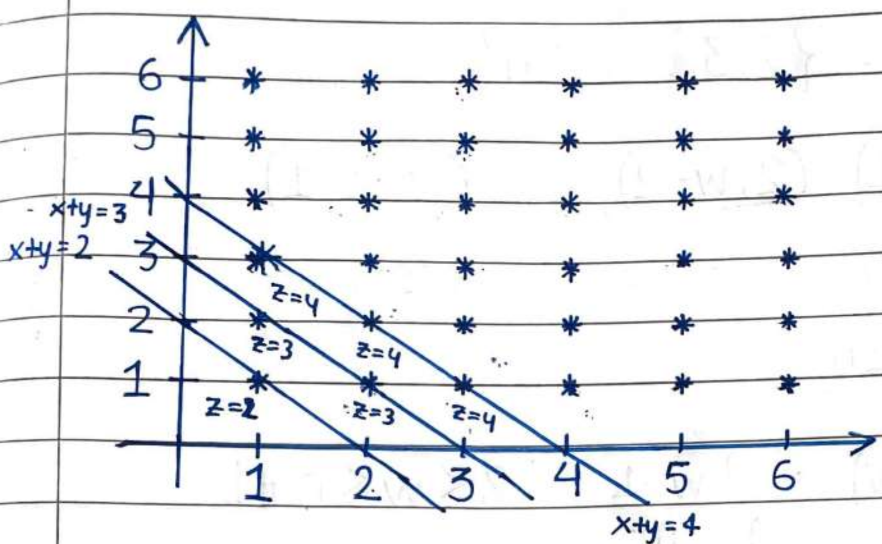
$X, Y \sim$ iid Unif$\{1,2,3,4,5,6\}$, $Z = X+Y$

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

classmate

Now how do we make contours?

① Picture the joint PMF on X-Y axis.



② Add over the contours:

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

Eg. Max, pair of dice

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

① Find range of Z : $Z \in \{1,2,3,4,5,6\}$

② Add over contours

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

Eg. iid Uniform $\{1, \ldots, n\}$: Sum

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

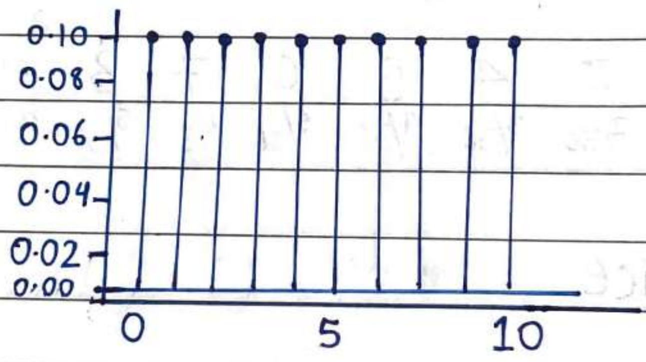
Step 1 : Range $\{2, 3, \ldots, 2n\}$

$W = w$: $(1, w-1), (2, w-2), \ldots, (w-1, 1)$

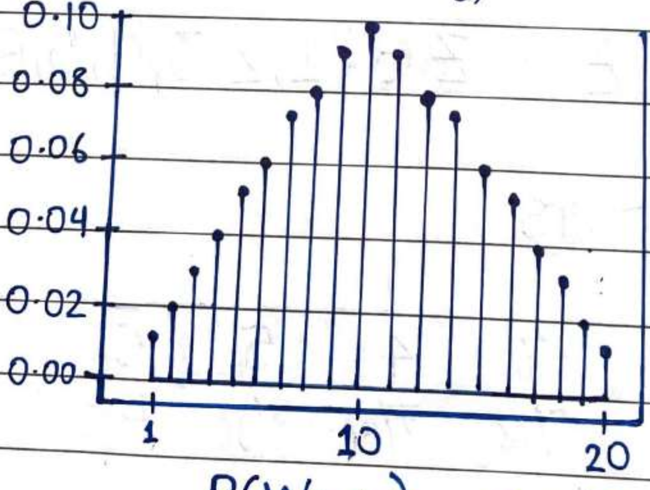
$\{w-1 \leq n\}$ $\quad$ $w-1 \leq n$

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

$$P(W = w) = \begin{cases} \dfrac{w-1}{n^2} & 2 \leq w \leq n+1 \\ \dfrac{2n-w+1}{n^2} & n+2 \leq w \leq 2n \end{cases}$$



P(X=x) and
also P(Y=y)



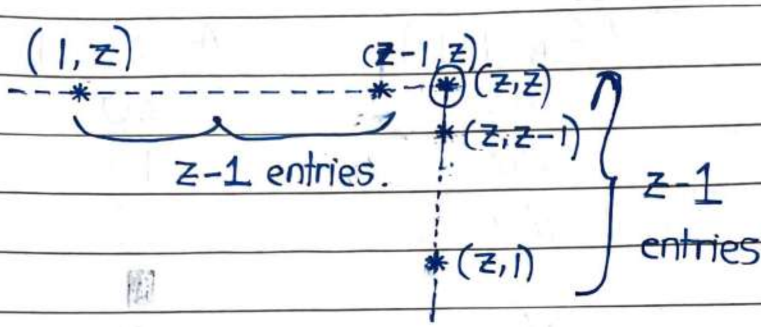
$P(W = w)$, $n = 10$
vs w

classmate

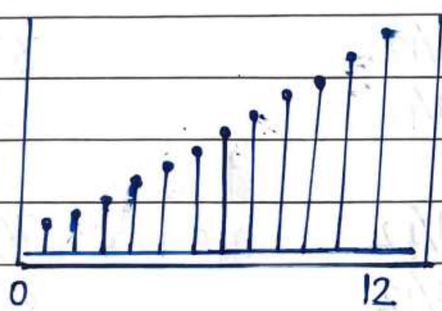
Eg. iid Uniform $\{1,2,\ldots,n\}$: Max

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

$Z \in \{1,\ldots,n\}$: $P(Z=z) = \dfrac{2z-1}{n^2}$



The individual X, Y PMFs are easy to visualise, same uniform stem plot in prev page.



$P(Z=z)$ vs z
$n = 12$.

★ <u>Functions of Multiple Random Variables</u>

Q. Fair die thrown twice. Pr(sum of 2 no's = 6)? PMF of sum?

Sum = $X_1 + X_2$ (fn of X_1 and X_2). $S \in \{2,3,4,\ldots,12\}$

$P(S=2) = \dfrac{1}{36}$ $P(S=3) = \dfrac{2}{36}$ ……

DATE

Q. Length of rectangle $L \sim \text{Uniform}\{5, 7, 9, 11\}$.
Given $L = l$, breadth $B \sim \text{Uniform}\{l-1, l-2, l-3\}$

f_{LB}

t_1	t_2	$f_{LB}(t_1, t_2)$	Area = LB
5	4	1/12	20
	3	1/12	15
	2	1/12	10
7	6	1/12	42
	5	1/12	35
	4	1/12	28
9	8	1/12	72
	7	1/12	63
	6	1/12	54
11	10	1/12	110
	9	1/12	99
	8	1/12	88

DATE

★ PMF of $g(X_1, X_2, \ldots, X_n)$

Suppose $X_1 \ldots X_n$ have joint PMF $f_{X_1 \ldots X_n}$ with T_{X_i} denoting range of X_i. Let $g : T_{X_1} \times T_{X_2} \times \ldots \times T_{X_n} \to \mathbb{R}$ be a fn. with range T_g. PMF of $X = g(X_1, X_2 \ldots X_n)$ is:→

$$\boxed{f_X(t) = P(g(X_1, X_2, \ldots, X_n) = t) = \sum_{(t_1, t_2, \ldots, t_n) : g(t_1, \ldots, t_n) = t} f_{X_1 \ldots X_n}(t_1, \ldots, t_n)}$$

Can be extended for joint PMF of 2 fns g and h.

Eg. Binomial from Bernoulli (p):

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

Eg. Sum of 2 random var. taking integer values, X and Y. Joint PMF f_{XY}, $Z = X + Y$

Let z be some integer.

$$P(Z = z) = P(X + Y = z)$$

$$= \sum_{x=-\infty}^{\infty} P(X = x, Y = z - x)$$

$$= \sum_{x=-\infty}^{\infty} f_{XY}(x, z - x)$$

$$= \sum_{y=-\infty}^{\infty} f_{XY}(z - y, y)$$

CONVOLUTION:

⊛ If X and Y are independent:

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

DATE

☆ Sum of 2 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 $X|Z$

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

$$= \frac{e^{-\lambda_1} e^{-\lambda_2}}{z!} \underbrace{\boxed{\sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda_1^x \lambda_2^{z-x}}}_{= (\lambda_1+\lambda_2)^z}$$

(Binomial formula of $(a+b)^n$)

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

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

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

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

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

$$= \boxed{\frac{n!}{k!(n-k)!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}}$$

$$X|Z \sim \text{Binomial}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$

$$Y|Z \sim \text{Binomial}\left(n, \frac{\lambda_2}{\lambda_1+\lambda_2}\right)$$

$X \sim Geo(p)$ $\quad Z = X - Y$
$Y \sim Geo(p)$
$f_X(j) = (1-p)^{j-1} p$

DATE

- If X and Y are independent, $g(X)$ and $h(Y)$ are independent for any 2 functions g and h.

- If X_1, X_2, X_3, X_4 are mutually independent,

  - $g(X_1, X_2)$ is independent of $h(X_3, X_4)$
  - $g(X_1, X_2, X_3)$ is independent of $h(X_4)$

- Functions of non overlapping sets of independent RV's are also independent.

1. Sum of independent Binomial (m, p) and Binomial (n, p)
2. Sum of independent Geometric (p) and Geometric (q)
3. Sum of r iid Geometric (p)
4. Sum of independent Neg-Binomial (r, p) and Neg-Binomial (s, p)

(1) If $X \sim \text{Binomial}(m, p)$, $Y \sim \text{Binomial}(n, p)$ $\quad X + Y \sim \text{Binomial}(m+n, p)$

(2) If $X \sim \text{Geometric}(p)$, $Y \sim \text{Geometric}(q)$ $\quad Z = X + Y$

$$\boxed{P(Z = n) = \frac{pq}{p-q}\left((1-q)^{n-1} - (1-p)^{n-1}\right) \quad n = 2, 3 \ldots}$$

(3) This sum $\underbrace{X_1 + X_2 + \ldots + X_n}_{S_n}$ has a negative binomial distribution with $P(S_n = m) = {}^{m-1}C_{n-1}\, p^n q^{m-n}$

(4) Sum $Z = X + Y$ $\quad Z \sim NB(r+s, p)$

DATE

✱ Minimum of 2 RV's

$X, Y \sim f_{XY}$

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

Eg. Throw a die ~~on~~ twice: Z = min of 2 no's seen.

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

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

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

For max, $= f_{XY}(z,z) + \sum_{t_2<z} f_{XY}(z,t_2) + \sum_{t_1<z} f_{XY}(t_1,z)$

✱ CDF (Cumulative distribution function) of maximum

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

$$F_X(x) = P(X \leq x)$$

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

$$F_Z(z) = P(\max(X,Y) \leq z)$$
$$= P((X \leq z) \text{ and } (Y \leq z))$$
$$= P(X \leq z)\, P(Y \leq z)$$
$$= F_X(z) F_Y(z)$$

CDF of maximum equal to ⇝ Product of CDFs.

DATE

Q. Let $X_1, \ldots, X_n \sim$ iid X. Find distribution

(1) $\max(X_1, \ldots, X_n)$

$$P(\max(X_1, \ldots, X_n) \leq z) = P(X_1 \leq z, X_2 \leq z, \ldots, X_n \leq z)$$
$$= (P(X \leq z))^n = (F_X(z))^n$$

(2) $$P(\min(X_1, \ldots, X_n) \geq z) = P(X_1 \geq z, X_2 \geq z, \ldots, X_n \geq z)$$
$$= (P(X \geq z))^n$$

Q. Let $X \sim$ Geometric(p), $Y \sim$ Geometric(p) be independent. Find dist. of $\min(X,Y)$.

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

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

$$P(\min(X,Y) = k) = P(\min(X,Y) \geq k) - P(\min(X,Y) \geq k+1)$$

$$= q^{k-1} - q^k = q^{k-1}(1-q)$$

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

i.e. $X_1 \sim$ Geometric(p_1) $X_2 \sim$ (Geometric(p_2)) Indep.

$$\min(X_1, X_2) \sim \text{Geometric}(1-(1-p_1)(1-p_2))$$
$$\sim \text{Geometric}(p_1 + p_2 - p_1 p_2)$$

Note. Try this for max. It will not be geometric.