08/12/19

Lecture 7 AMS 131

X, Y: random variables, the disjoint or bivariate distribution of (X, Y) is the collection $P[(X,Y) \in C]$ of all probabilities for all sets $C \subset \mathbb{R}^2$ such that $(X, Y) \in C$ isn't weird

1. X and Y are both discrete

If there are only finitely or countably infinite many possible values (x,y) – X and Y have a discrete joint distribution

The joint probability mass function of (X,Y) discrete is $f_{X,Y}(x,y) = P(X = x, Y = y)$

$\{(x,y) : f_{XY}(x,y) > 0\}$ is the support of f_{XY}

- $\sum f_{XY}(x,y) = 1$
- $P[(X,Y) \in C] = \sum_{(x,y) \in C} f_{XY}(x,y)$

2. X and Y are continuous

If you can find a non negative function $f_{XY}(x,y)$ defined for all $(x,y) \in \mathbb{R}^2$ such that for every non weird subset C:

$$P(X, Y \in C) = \iint_C f_{XY}(x,y)\,dx\,dy$$

$\{(x,y) : f_{XY}(x,y) > 0\}$: support of X, Y dist

- For all (x,y) in $\mathbb{R}^2$, $f_{XY}(x,y) \geq 0$, and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$
- If (X,Y) have a continuous joint distribution then X and Y have a continuous univariate marginal distribution when separate
- For all continuous pdfs $f_{XY}(x,y)$: every individual point and every countable infinite set/sequence has probability 0

AMS 131 | Lecture 7 (cont.)

- If g is a continuous function of one real variable defined on (a,b) then $\{(x,y): y=g(x), a<x<b\}$ and $\{(x,y): x=g(y), a<y<b\}$ also have probability 0 (3D: curve as probability 0)

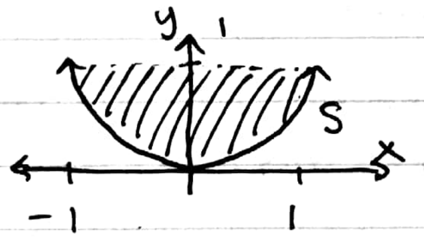
This means the converse of the univariate statement is not true. If X has a continuous distribution on $\mathbb{R}=\mathbb{R}'$ and $Y \triangleq X$, then both X and Y are continuous, but:

$$P\left[\begin{matrix}(X,Y) \text{ lies on} \\ \text{the line } y=x\end{matrix}\right]=1$$

so (X,Y) can't have a continuous joint distribution

$$\int_0^1\int_0^1 f_{XY}(x,y)\,dx\,dy = 1$$

$$f_{XY}(x,y) = \begin{cases} cx^2y & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$ joint PDF

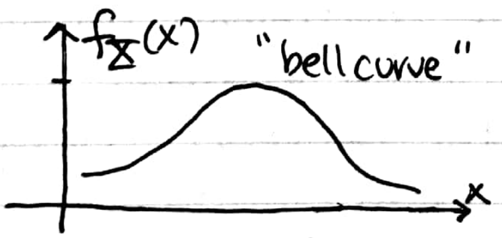


$x^2: [0,1]$, $y: [0,1]$ then look at $y \ge x^2$

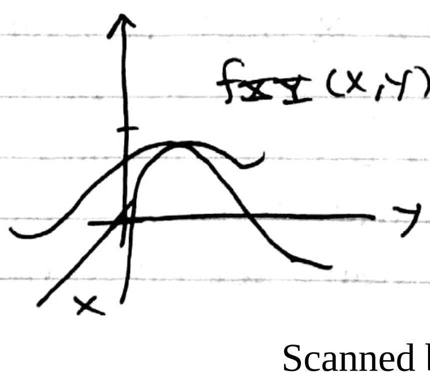
$$1 = \iint_S cx^2y\,dx\,dy$$

- strictly ≥ 0 probability
- double integral over whole plane $= 1$

ex. of two R.V. PDF



1 variable, 2D



"3D bell curve"

2 variables 3D

$$c\left[\int_{-1}^{1}\left[\int_{x^2}^{1}(x^2y)\,dy\right]dx\right]$$

x goes from $[-1,1]$
y bottom is x^2, top 1

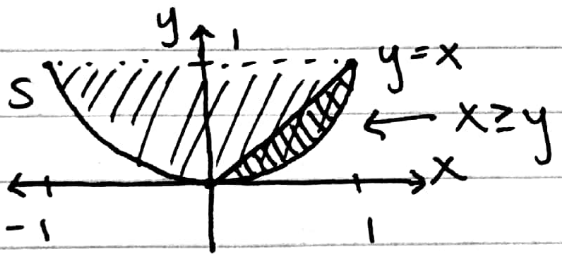
* practice finding : limits of integration

$$c\left[\int_{0}^{1}\left[\int_{-\sqrt{y}}^{+\sqrt{y}}(x^2y)\,dx\right]dy\right]$$

y goes from $[0,1]$
x: $y = x^2$, $x = \pm\sqrt{y}$

- both orders of integrations are equal $= \frac{4c}{21} = 1$
- Fix outside integral with constants $c = \frac{21}{4}$
  vary inside integral based on ↑

Now let's see $P(X \geq Y)$



- find where $x=y$
- then, see where $x \geq y$

$$P(X \geq Y) = \iint_S f_{XY}(x,y)\,dy\,dx$$

$$\frac{21}{4}\int_0^1\left[\int_{x^2}^{x} x^2y\,dy\right]dx$$

- area is only $x>0$
- y is between x^2, x

$$\frac{21}{4}\int_0^1\left[\int_{y}^{\sqrt{y}} x^2y\,dx\right]dy$$

- y $[0,1]$
- between $y=x$ $x=y$
  $y=x^2$ $x=\sqrt{y}$

$$P(X \geq Y) = \frac{3}{20}$$

AMS 131 Lecture 7 (cont.)

3. Mixed bivariate distribution: one of each

X is discrete, Y continuous

$P(X \in A$ and $Y \in B)$ subsets of $\mathbb{R}$

$$= \int_B \sum_{x \in A} f_{XY}(x,y)\,dy$$

f_{XY} is the joint pmf / pdf.

If X is $x_1, x_2 \ldots$ then: $\int_{-\infty}^{\infty} \sum_{i=1}^{\infty} f_{XY}(x_i, y)\,dy = 1$

T: patients get treatment

C: placebo / best treatment

$$X_i = \begin{cases} 1 & \text{if patient } i \text{ in } T \text{ is success} \\ 0 & \text{else} \end{cases}$$

and θ = proportion of patients who would have no relapse if they were in the study $(0,1)$ and (X_i, θ) has mixed bi-distr.

$(X \mid \theta) \sim \text{Bernoulli}(\theta)$ and $\theta \sim \text{Uniform}(0,1)$

$$f_{X,\theta}(x,\theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & x = 0,1 \quad 0 < \theta < 1 \\ 0 & \text{else} \end{cases}$$

ex: $P(X=1) = P(X=1$ and θ between $0,1)$

$$= \int_0^1 \theta^1 (1-\theta)^{1-1}\,d\theta = \int_0^1 \theta\,d\theta = \frac{1}{2}$$

Lecture 7 (cont.) AMS 131

The joint CDF is $F_{XY}(x,y) = P(X \le x \text{ and } Y \le y)$

1. You can obtain the marginal CDF

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x,y) \quad \text{and} \quad F_Y(y) = \lim_{x \to \infty} F_{XY}(x,y)$$

Covers all possibilities for x/y and all that's left is y/x.

2. PDF $\leftrightarrow$ CDF

$$F_{XY} = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(r,s)\, dr\, ds$$

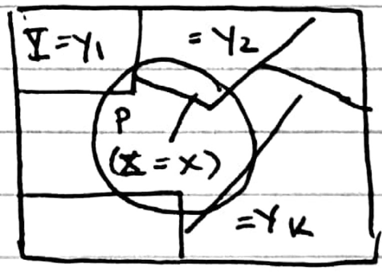
$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$$ partial derivatives

3. If (X,Y) has a discrete joint dist. with joint pmf $f_{XY}(x,y)$, then the marginal pmf

$$f_X(x) = \sum_y f_{XY}(x,y)$$

$$f_{XY}(x,y) = P(X = x \text{ and } Y = y)$$

$$f_X(x) = P(X = x \text{ and } Y = y_1) + P(X = x \text{ and } Y = y_2) \dots + P(X = x \text{ and } Y = y_K)$$



to get marginal pdfs of x, sum over y or take integral of y

$$f_X(x) = \sum_{\text{all } y} f_{XY}(x,y)$$ marginalizing over Y

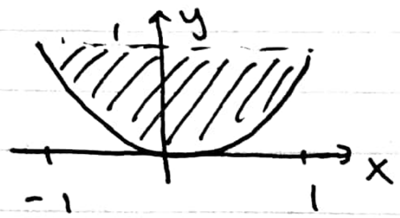
$\underset{\sim}{X} = (X_1, X_2 \dots X_n)$ continuous

$$f_{X_1}(x) = \int_{x_2} \int_{x_3} \dots \int_{x_n} f_X(x)\, dx_2 \dots dx_n$$

AMS 131 — Lecture 7 (cont.)

marginal PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy \qquad \text{and vice versa}$$



$-1 \le X \le 1 \qquad (-1, 1)$

marginal PDF $\quad f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$

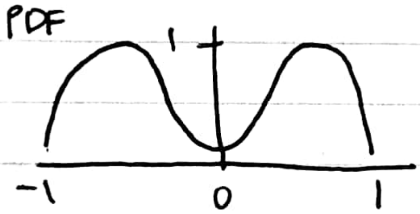
· fix X, integrate over y

$$= \int_{x^2}^{1} \frac{21}{4} x^2 y\,dy = \frac{21}{8} x^2(1-x^4) \qquad -1 < x < 1$$

0 else

PDF

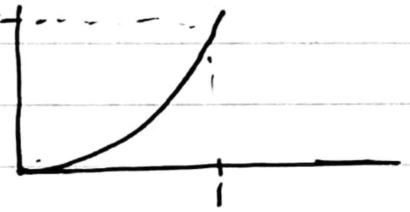
$f_X(x)$



$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

$$= \int_{-\sqrt{y}}^{+\sqrt{y}} \frac{21}{4} x^2 y\,dx = \begin{cases} \frac{7}{2} y^{\frac{5}{2}} & 0 < y < 1 \\ 0 & \text{else} \end{cases}$$

$f_Y(y)$ 3.5



If you have the joint dist. $f_{XY}(x,y)$ you can reconstruct the marginals $f_X(x)$ and $f_Y(y)$, but not the other way around - they do not uniquely determine the joint.

Harder to visualize a joint distribution (2-dimensional) versus each of its 1-dimensional marginal distributions

Lecture 7 (cont.) AMS 131

Case 1: X = # of heads in n coin toss 1
Y = # of heads in n coin toss 2

Case 2: X = # of heads in n coin toss 1
$Y = X$

1. $X \sim$ Binomial $(n, \frac{1}{2})$
so $f_X(x) = \begin{cases} \binom{n}{x}(\frac{1}{2})^x(1-\frac{1}{2})^{n-x} & x = 0,1,\dots n \\ 0 & \text{else} \end{cases}$

$Y \sim$ Binomial $(n, \frac{1}{2})$
so $f_Y(y) = \begin{cases} \binom{n}{y}(\frac{1}{2})^n & y = 0,1,\dots n \\ 0 & \text{else} \end{cases}$

since X and Y are independent,
$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

$$f_{XY}(x,y) = \begin{cases} \binom{n}{x}\binom{n}{y}(\frac{1}{2})^{2n} & x = 0,1\dots n \quad y = 0,1\dots n \\ 0 & \text{else} \end{cases}$$

2. X is binomial $(n, \frac{1}{2})$ and so is Y but their joint distribution (since $Y = X$) is:

$$f_{XY}(x,y) = \begin{cases} \binom{n}{x}(\frac{1}{2})^n & \text{for } x = y = 0,1\dots n \\ 0 & \text{else} \end{cases}$$

their joint distributions are the same instead of combined

Marginals can determine, uniquely, the joint when X and Y are independent

X and Y are independent if for every (non-weird) sets of A and B of real numbers

$$P(X \in A \text{ and } Y \in B) = P(X \in A) \cdot P(Y \in B)$$

- $F_{XY}(x,y) = P(X \leq x \text{ and } Y \leq y) = P(X \leq x) P(Y \leq y) = F_X(x) \cdot F_Y(y)$

differentiate with respect to x and y

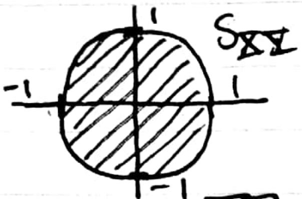
X, Y independent $\overset{\text{iff}}{\longleftrightarrow}$ $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

Scanned by CamScanner

AMS 131

Lecture 7 (cont.)

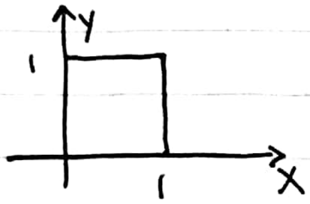
$f_{XY}(x,y) = \begin{cases} Kx^2y^2 & \text{for } 0 \le x^2+y^2 \le 1 \\ 0 & \text{else} \end{cases}$ joint PDF



S_{XY} support set

$\iint_S Kx^2y^2\,dx\,dy = 1$ normalizing constant

$1 = \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} Kx^2y^2\,dx\,dy = \frac{\pi}{24}$ $K = \frac{24}{\pi}$ (not independent $\underline{x^2+y^2} \le 1$)



$X \sim U(0,1)$
$Y \sim U(0,1)$ $f_{XY}(x,y) = \begin{cases} 1 & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$

good chance of independence not linked to

Continuous R.V. X and Y have joint PDF:

$f_{XY}(x,y) = \begin{cases} Ke^{-(x+2y)} & x \ge 0,\ y \ge 0 \\ 0 & \text{else} \end{cases}$

chance of independence based on support, no linkage

x and y are independent because $e^{-(x+2y)}$ can factor into $(e^{-x})(e^{-2y})$ and the support factors

$\iint_S Ke^{-(x+2y)}\,dx\,dy = 1 \rightarrow \int_0^\infty K_x e^{-x}\,dx = 1,\ \int_0^\infty K_y e^{-2y}\,dy = 1$ $K = K_x \cdot K_y$, $K_x = 1, K_y = 2$

Independence: for AND and GIVEN

$\underset{\sim}{X} = (X_1, \dots, X_n)$ $f_X(x) = ?$

$f_{X_1 \dots X_n}(x_1 \dots x_n) = P(X_1 = x_1, \overset{\text{and}}{\dots} X_n = x_n)$ not independent

$P(X_1 = 1)P(X_2 = x_2 | X_1 = 1)P(X_3 = x_3 | X_1 = x_1, X_2 = x_2) =$ mess

if independent, joint = product of marginals

Conditional probability $P(B|A) = \frac{P(A \cap B)}{P(A)}$

$P(X = x | Y = y)$ discrete

If X and Y are discrete with a joint distribution pmf $f_{XY}(x,y)$ and X has marginal pmf $f_X(x)$ then for each x, $f_X(x) > 0$, then $\rightarrow$

Lecture 7 (cont.) AMS 131

$$f_{Y|X}(y|x) \triangleq \frac{f_{XY}(x,y)}{f_X(x)} = P(Y=y \mid X=x)$$

the conditional pmf of Y given X

Gender and MLP $X = \begin{cases} 1 & \text{if Female} \\ 0 & \text{Male} \end{cases}$ $Y = \begin{cases} 1 & \text{if Yes MLP} \\ 0 & \text{no} \end{cases}$

$P(Y=1 \mid X=1) = P(Y \mid F) = \frac{29}{49}$

	MLP Y ($Y=1$)	N ($Y=0$)	
($X=1$) F	29	20	49
($X=0$) M	52	5	57
	81	25	106

$$= \frac{P(Y=1 \text{ and } X=1)}{P(X=1)} = \frac{\frac{29}{106}}{\frac{49}{106}} = \frac{29}{46}$$

works for distribution of R.V.

If X and Y have a continuous joint distribution with PDF $f_{XY}(x,y)$ and X has marginal (continuous) PDF $f_X(x)$, then for each x, $f_X(x) > 0$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$ conditional PDF of Y given X

X, Y have joint PDF $f_{XY}(x,y) = \begin{cases} \frac{21}{4}x^2 y & 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$

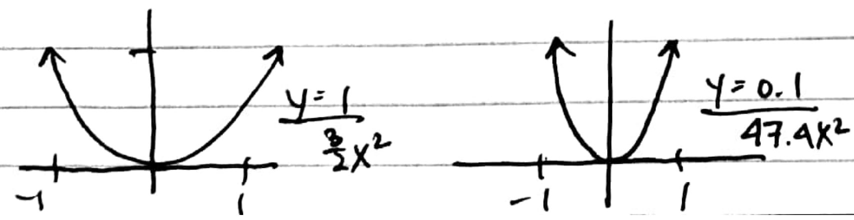
marginals

$$f_X(x) = \begin{cases} \frac{21}{8}x^2(1-x^4) & -1 \le x \le 1 \\ 0 & \text{else} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{7}{2}y^{\frac{5}{2}} & 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1-x^4)} \text{ for } 0 \le x^2 \le y \le 1 \quad 0 \text{ else} = \frac{2y}{1-x^4} \quad 0 \text{ else}$$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{\frac{21}{4}x^2 y}{\frac{7}{2}y^{\frac{5}{2}}} = \begin{cases} \frac{3}{2}x^2 y^{-\frac{3}{2}} & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$



$f_{X|Y}(x|y)$

AMS 131

Lecture 7 (cont.)

When X and Y are continuous, computing $f_{Y|X}(y|x)$ relies on conditioning $X=x$ which has a probability of 0, but:

$$f_{Y|X}(y|x^*) = \lim_{\epsilon \downarrow 0} \frac{\partial}{\partial y} P(Y \le y \mid x^* - \frac{\epsilon}{2} \le X \le x^* + \frac{\epsilon}{2})$$

Take a strip of x values of width ϵ and around $X = x^* \rightarrow P(Y=y \mid X$ is in the strip), differentiate with respect to y and let ϵ go to 0. Thus,

$f_{Y|X}(y|x)$ is like the conditional PDF of Y given that X is <u>close</u> to x.

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \quad \text{and} \quad f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$\rightarrow f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y)$$

Connecting a joint PDF with a marginal and conditional PDF

A machine produces nuts and bolts. - the nut can be too big/small, bolt too big/small

A nut/bolt pair is defective if they don't fit.

θ = proportion of defective bolts, over long time

↳ constant over time ∵ stationarity

We can only observe over a short time interval,

θ is unknown - unknown, truth, data

- Take random sample of m $\quad D_i = \begin{cases} 1 & \text{defective} \\ 0 & \text{else} \end{cases}$
- # of defectives N

$(D_i \mid \theta) \sim \text{Bernoulli}(\theta) \quad i = 1, \ldots, m \quad$ IID, stationarity

vs. $D_i \sim \text{Bernoulli}(\theta)$ first 10 help decide θ

which becomes conditionally independent given θ

$N = \sum_{i=1}^{m} D_i \quad f_N(n|m,\theta) =$ sampling distribution $\quad \hat{\theta} = \frac{N}{m} = \frac{3}{114} = 2.6\%$

an estimate of θ

Lecture 7 (cont.) AMS 131

Bayesian story θ = unknown

$D = (D_1, \ldots, D_m)$ vector

probability	statistics HARD
P(data | unknown)	P(unknown | data)
$P(N \mid \theta)$ EASY	$P(\theta \mid D)$ stat. inference
$P(D \mid \theta) = P(\theta \mid D)$	$= P(\theta \mid N)$ because
↳ Bayes Theorem	$\underset{\sim}{D} = (D_1, \ldots, D_m)$ and N
	carry same info about θ

$$P(\theta \mid D) = \frac{P(\theta) P(D \mid \theta)}{P(D)}$$

$$\updownarrow$$

$$P(\theta \mid N) = \frac{P(\theta) P(N \mid \theta)}{P(N)}$$

supposed to be lowercase P, Bayesian story = functions

total info about θ

info about θ external to dataset

info about θ internal to the dataset

normalizing constant

Multivariate distributions: finite # of R.V.

$Y_1, \ldots, Y_n$ n = positive, finite integer

The joint CDF of n R.V.s is the function:

- $F_{Y_1, \ldots, Y_n}(y_1 \cdots y_n) = P(Y_1 \le y_1, \overset{\text{and}}{\ldots}, Y_n \le y_n)$
- $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$, $\underset{\sim}{y} = (y_1, \ldots, y_n)$ vector notation
- $F_{\underset{\sim}{Y}}(\underset{\sim}{y}) = P(Y_1 \le y_1, \ldots, Y_n \le y_n)$

$\underset{\sim}{Y}$ is a random vector taking place in $\mathbb{R}^n$

$\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ have a discrete joint distribution if the random vector $\underset{\sim}{Y}$ can only be a finite or countably infinite # of possible values $(y_1, \ldots, y_n) \in \mathbb{R}^n$.

The joint PMF of $\underset{\sim}{Y}$ is $f_{Y_1, \ldots Y_n}(y_1 \ldots y_n) = P(Y_1 = y_1 \ldots Y_n = y_n)$

$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = P(\underset{\sim}{Y} = \underset{\sim}{y})$

n patients $B_i = \begin{cases} 1 & \text{good outcome} \\ 0 & \text{else} \end{cases}$ if nothing else is known about the patients:

model $B_i \sim$ Bernoulli(θ) IID $\underset{\sim}{B} = (B_1, \ldots, B_n)$ $\underset{\sim}{b} = (b_1 \ldots b_n)$

$f_{\underset{\sim}{B}}(\underset{\sim}{b}) = P(B_1 = b_1 \ldots B_n = b_n)$, if you know θ, use $f_{\underset{\sim}{B}}(\underset{\sim}{b})$ to predict!

AMS 131

Lecture 7 (cont.)

$P(B_1=b_1, \ldots B_n=b_n) = P(B_1=b_1) \ldots P(B_n=b_n)$ multiply

n. R.V. $Y_1 \ldots Y_n$ have a continuous joint distribution if you can find a function $f_{\underline{Y}}$ on $\mathbb{R}^n$ such that for every non weird subset $C \in \mathbb{R}^n$

$$P[(Y_1, \ldots, Y_n) \in C] = \int_C \cdots \int f_{Y_1 \ldots Y_n}(y_1 \ldots y_n)\, dy_1 \ldots dy_n$$

$f_{\underline{Y}}(\underline{y})$ is the joint PDF of $\underline{Y}$

$$P(\underline{Y} \in C) = \int_C \cdots \int f_{\underline{Y}}(\underline{y})\, d\underline{y}$$

If the joint dist. of $\underline{Y}$ is continuous then

$$f_{\underline{Y}}(\underline{y}) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{\underline{Y}}(\underline{y})$$

Mixed, discrete, continuous with n R.V. Continuous random vectors behave the same as 2 R.V.

$(\underline{B}, \theta) = (B_1 \ldots B_n, \theta)$ B_i = discrete, $0 < \theta < 1$ = continuous