ch 1-2 reviews a lot of material you should know.

ch1 1]* p2 The sample space S is the set of all possible outcomes of an experiment.

2]* p2 An event A is a subset of $\mathcal{B} \subseteq S$

3]* p2 Events $A_1, A_2, \ldots$ are pairwise <u>disjoint</u> or <u>mutually exclusive</u> if $A_i \cap A_j = \phi$ for $i \neq j$.

4] p2 A <u>probability function</u> is a set function $P: \mathcal{B} \to [0,1]$ where $\mathcal{B}$ is a special field of subsets of S ($\mathcal{B}$ is the smallest σ-field containing all open sets for pdfs).

ex] Flip coin $S = \{H, T\}$ $P(\{H\}) = \frac{1}{3}$
$P(\{T\}) = \frac{2}{3}$. Will often write $P(H) = \frac{1}{3}$

5] p3 Th 1.2 $P(\phi)$

$P(A)$

$P(A^c)$

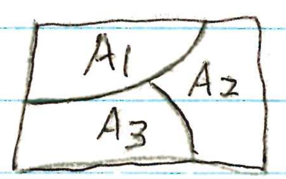
$P(A \cup B)$

$A \subseteq B$

6] p3 Boole's and Bonferonni's inequalities give useful bounds.

7] p6 If $A_1, \ldots$ are pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = S$, then $A_1, \ldots,$ is a partition of S.

ex]



S

8]* $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

9)* p5 $A, B \in S$, $P(B>0)$ **Conditional prob**

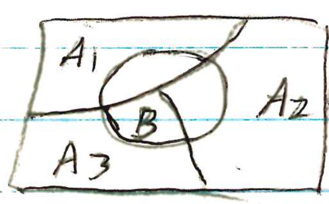
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

10]* p5 A and B are **independent** if $P(A \cap B) = P(A)P(B)$
written $A \perp\!\!\!\perp B$.
If $A \perp\!\!\!\perp B$, then $A \perp\!\!\!\perp B^c$, $A^c \perp\!\!\!\perp B$, $A^c \perp\!\!\!\perp B^c$.

11]* p5 A collection of events $A_1, \dots, A_n$ are mutually independent if for any subcollection $A_{i1}, \dots, A_{ik}$, $P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$.

12] p6 Bayes Theorem Let $A_1, \dots$ be a partition and $P(B)>0$. Then

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)} \quad \begin{matrix} \leftarrow P(A_i \cap B) \\ \leftarrow P(B) \end{matrix}$$



13]* p7 A random variable $Y: S \to \mathbb{R}$

14] p7 $P_S[\{s \in S : Y(s) = y\}] = \underbrace{P_Y(Y = y_i)}_{\text{induced prob function}}$

usually omit S.

The range of Y is the sample space $\mathcal{Y}$ of Y and $P_Y : \mathcal{Y} \to [0,1]$.

ex] $S = \{HH, HT, TH, TT\}$
$P(s_j) = \frac{1}{4}$ $j = 1, \dots, 4$ $Y = \#$ heads

y	0	1	2
$P(Y=y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

15]* p8 The cdf of Y is $F_Y(y) = P(Y \le y)$, $y \in \mathbb{R}$
$F(-\infty) = 0$, $F(\infty) = 1$, F is nondecreasing and right continuous.

16]* p8 Y is continuous if F is (absolutely continuous)
Y is discrete if F is a step function

17] * p8 The pmf of discrete Y is

$$f_Y(y) = f(y) = P(Y = y) \quad \forall y \in \mathbb{R}.$$

But often take support $\mathcal{Y} = \{y \mid f(y) > 0\}$ and define $f(y)$ $\forall y \in \mathcal{Y}$.

18] * p8 The pdf of continuous Y satisfies

$$F(y) = \int_{-\infty}^{y} f(t)\,dt \quad \forall y \in \mathbb{R}.$$

Note that $\frac{d}{dy} F(y) = F'(y) = f(y)$

19) f is a pdf iff $f(y) \geq 0$ $\forall y$ and $\int_{-\infty}^{\infty} f(y)\,dy = 1$

" pmf $\sum_{y} f(y) = 1$ "

20] * p9
$$E\, g(Y) = \begin{cases} \int_{-\infty}^{\infty} g(y) f(y)\,dy & Y \text{ contin} \\ \sum_{y \in \mathcal{Y}} g(y) f(y) & Y \text{ disc} \end{cases}$$

provided the integral or sum exists if $g(y)$ is replaced by $|g(y)|$. If $E|g(Y)| = \infty$, $E\,g(Y)$ does not exist.

21] In particular
$$E\,Y = \begin{cases} \int_{-\infty}^{\infty} y f(y)\,dy & Y \text{ contin} \\ \sum_{y \in \mathcal{Y}} y f(y) & Y \text{ disc} \end{cases}$$

22] * p9 $VAR(Y) = V(Y) = E(Y - E(Y))^2$

$SD(Y) = \sqrt{V(Y)}$

23] p9 Know $V(Y) = E(Y^2) - [E(Y)]^2$

24] p10 $V(aY + b) = a^2 V(Y)$

25] * p10 The mgf of Y is $m(t) = E\,e^{tY}$ if $\exists\, t_0 \ni E\,e^{tY}$ exists $\forall\, t \in (-t_0, t_0)$.

So $m_Y(t) = \int_{-\infty}^{\infty} e^{ty} f(y)\,dy$ $\quad$ Y contin $\quad$ 2.5

$\sum_{y \in \mathcal{Y}} e^{ty} f(y)$ $\quad$ Y disc

26]* p11 X and Y are identically distributed, $X \sim Y$ or $Y \sim F_X$ if $F_X(y) = F_Y(y)$ $\forall y$.

27] p11 The kth moment of Y is $E\,Y^k$

the kth central moment of Y is $E\left[(Y - E(Y))^k\right]$

28] p11 Th 1.13 Let $m^{(k)}(t) = \frac{d^k}{dt^k} m(t)$.

Then $E(Y^k) = m^{(k)}(0)$.

§1.5 The kernel method

29]* Let $f(y \mid \underline{\theta}) = \underbrace{c(\underline{\theta})}_{\text{constant}} \underbrace{k(y \mid \underline{\theta})}_{\text{kernel = "essential part"}}$.

where $\underline{\theta} = (\theta_1, \dots, \theta_J)$ are known parameters

for $\underline{\theta} \in$ parameter space Θ.

ex] $Y \sim N(\mu, \sigma^2)$, $\underline{\theta} = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times (0, \infty)\}$
$= \{(\mu, \sigma) \mid \mu \in \mathbb{R}, \sigma > 0\}$

Then $1 = \int_{-\infty}^{\infty} c(\underline{\theta})\, k(y \mid \underline{\theta})\, dy$ so

$\frac{1}{c(\underline{\theta})} = \int_{-\infty}^{\infty} k(y \mid \underline{\theta})\, dy$.

Often $E\,g(Y) = \int_{-\infty}^{\infty} g(y)\, f_Y(y \mid \underline{\theta})\, dy$

$= a\, c(\underline{\theta}) \underbrace{\int_{-\infty}^{\infty} k(y \mid \underline{\tau})\, dy}_{\frac{1}{c(\underline{\tau})}} = \frac{a\, c(\underline{\theta})}{c(\underline{\tau})} \underbrace{\int_{-\infty}^{\infty} c(\underline{\tau})\, k(y \mid \underline{\tau})\, dy}_{1 = \int_{-\infty}^{\infty} f_Y(y \mid \underline{\tau})\, dy}$

or $E\, g(Y) = \frac{a\, c(\theta)}{c(\tau)}$ where $(\tau) \in \Theta$.

Replace integral by sum for discrete RVs.

This trick is often used for Beta Gamma and normal distributions.

30] Another common trick is "complete the square" for the normal dist.

$A = e^{\frac{-1}{2\sigma^2}(y^2 - 2a\mu y + b)}$ has

$\frac{-1}{2\sigma^2}(y^2 - 2a\mu y + b) =$

$-\frac{1}{2\sigma^2}(y^2 - 2a\mu y + \overbrace{a^2\mu^2 - a^2\mu^2}^{0} + b) =$

$\frac{-1}{2\sigma^2}(y - a\mu)^2 \underbrace{\frac{-1}{2\sigma^2}(-a^2\mu + b)}_{c}$

so $A = \underbrace{e^{-\frac{1}{2\sigma^2}(y - a\mu)^2}}_{N(a\mu,\, \sigma^2) \text{ kernel}} \underbrace{e^c}_{d = e^c}$

See ex 1.11 - 1.13.

31] Common exam 1 problem Find $E\, g(Y)$ or $\int k(y|\tau)\, dy = \frac{1}{c(\tau)}$ using the kernel method.

ex] Let $f(y) = \frac{1}{\lambda} y^{\frac{1}{\lambda} - 1}$, $y \in (0,1)$, $\lambda > 0$.

a) $E\, Y^k = \int_0^1 y^k \frac{1}{\lambda} y^{\frac{1}{\lambda} - 1}\, dy =$

$$\frac{1}{\lambda}\int_0^1 y^{k+\frac{1}{\lambda}-1}\,dy = \frac{1}{\lambda}\int_0^1 y^{\frac{1}{\tau}-1}\,dy$$

$$\frac{1}{\tau} = k+\frac{1}{\lambda} = \frac{k\lambda+1}{\lambda} \Rightarrow \tau = \frac{\lambda}{k\lambda+1}$$

$$= \frac{\tau}{\lambda}\underbrace{\int_0^1 \frac{1}{\tau}\, y^{\frac{1}{\tau}-1}\,dy}_{1=\int f(y|\tau)dy} = \frac{\tau}{\lambda} = \frac{1}{k\lambda+1}$$

if $\tau > 0$ or $k\lambda+1>0$ or $k > \frac{-1}{\lambda}$.

b) $\int_0^1 y^{\frac{1}{\tau}-1}\,dy = \tau$ if $\tau>0$ since $\int_0^1 \frac{1}{\lambda}\, y^{\frac{1}{\lambda}-1}\,dy = 1$.

32] *p. 11 Let $f(y) = f(y|\underline{\theta})$ be the pdf or pmf. The support of Y is

$\mathcal{Y}_{\underline{\theta}} = \{y \mid f(y|\underline{\theta}) > 0\}$. We use $\mathcal{Y} = \{y \mid f(y|\underline{\theta}) > 0\}$ if the support

does not depend on $\underline{\theta} \in \Theta$. The parameter space Θ is the set of parameters of interest.

33] *p12 The indicator function $I_A(x) \equiv I(x\in A)$ $= \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$. Often $I_{(0,\infty)}(y)$ is

denoted by $I(y>0)$.

34] Suppose $E\,Y^k = E\,X^k$ for $k = 1, 2, \ldots$. X and Y do not necessarily have the same distribution and the mgf of X may not exist (ie if X is lognormal).

35) prop 1.12 d It $m_X(t) = m_Y(t)$ $\forall\, t \in (-t_0, t_0)$ then $X \sim Y$.

36] Common exam 1 problem p 13-14 ex 1.8

Given $m(t)$ find $EY = m'(0)$, $EY^2 = m''(0)$ and $V(Y) = EY^2 - [E(Y)]^2$.

ex] Suppose $P(Y=y) = \rho(1-\rho)^{y-1}$, $y = 1, 2, \ldots$ and $0 \le \rho \le 1$ ($Y = X+1$ where $X \sim geom(\rho)$).

$$m(t) = \sum_{y=1}^{\infty} e^{ty} \rho(1-\rho)^{y-1} = \frac{1}{1-\rho} \sum_{y=1}^{\infty} e^{ty} \rho(1-\rho)^{y}$$

$$= \frac{\rho}{1-\rho} \sum_{y=1}^{\infty} \left[e^t(1-\rho)\right]^y$$

Geometric series p 292 (beginning of ch 10)

$$\sum_{y=y_1}^{\infty} a^y = \frac{a^{y_1}}{1-a} \quad \text{if } |a|<1 \text{ and } y_1 \ge 0$$

So $m(t) = \frac{\rho}{1-\rho} \frac{e^t(1-\rho)}{1-e^t(1-\rho)} = \frac{\rho e^t}{1-(1-\rho)e^t}$, for $t < -\log(1-\rho)$ since $0 \le a = e^t(1-\rho) < 1$ if $e^t < \frac{1}{1-\rho}$ or $t < \log \frac{1}{1-\rho}$.

Now $m'(t) = \frac{[1-(1-\rho)e^t]\rho e^t + \rho e^t[(1-\rho)e^t]}{[1-(1-\rho)e^t]^2} = \frac{\rho e^t}{[1-(1-\rho)e^t]^2}$

quotient rule $\frac{dn' - nd'}{[d]^2}$

So $EY = m'(0) = \frac{\rho}{[1-(1-\rho)]^2} = \boxed{\frac{1}{\rho}}$

$$m''(t) = \frac{[1-(1-\rho)e^t]^2 \rho e^t - \rho e^t\, 2[1-(1-\rho)e^t][-(1-\rho)e^t]}{[1-(1-\rho)e^t]^4}$$

$$= \frac{\rho e^t [1-(1-\rho)e^t + 2(1-\rho)e^t]}{\ldots}$$

so $EY^2 = m_Y''(0) = \dfrac{\rho\,[1-(1-\rho)+2(1-\rho)]}{[1-(1-\rho)]^3} =$

$\dfrac{\rho\,(1+1-\rho)}{[1-(1-\rho)]^3} = \dfrac{2-\rho}{\rho^2}$ and

$V(Y) = EY^2 - (EY)^2 = \dfrac{2-\rho}{\rho^2} - \dfrac{1}{\rho^2} = \dfrac{1-\rho}{\rho^2}.$

37] The Kernel method is a summation or integration technique like u substitution or integration by parts.

ex] $Y \sim \text{poisson}(\theta)$ $\quad f(y) = \dfrac{e^{-\theta}\theta^y}{y!}, \theta>0, y=0,1,\ldots$

$\underbrace{\text{so } \sum_{y=0}^{\infty} \dfrac{\tau^y}{y!} = e^{\tau},}_{\text{useful for HW1}}$ $\tau>0$ since $\sum_{y=0}^{\infty} \dfrac{e^{-\theta}\theta^y}{y!} = 1.$

ex] $Y \sim \text{gamma}(\nu,\lambda)$ $\quad f(y) = \dfrac{y^{\nu-1} e^{-y/\lambda}}{\lambda^{\nu}\Gamma(\nu)}, y,\nu,\lambda>0$

Take $\lambda=1$. $\int_0^{\infty} y^{\nu-1} e^{-y}\,dy = \Gamma(\nu).$

Fact p292 $\Gamma(x+1) = x\Gamma(x)$ is useful for HW1.

ex] (logarithmic(θ)) Suppose $f(x) = \dfrac{-1}{\log(1-\theta)} \dfrac{\theta^x}{x}, x=1,2,\ldots, 0<\theta<1$

Find $m(t)$. $\quad \log(x) = \ln(x)$

soln] $m(t) = E e^{tX} = \sum_{x=1}^{\infty} e^{tx}\left(\dfrac{-1}{\log(1-\theta)}\right)\dfrac{\theta^x}{x} =$

$\sum_{x=1}^{\infty} \dfrac{-1}{\log(1-\theta)} \dfrac{(\theta e^t)^x}{x} = \dfrac{\log(1-\theta e^t)}{\log(1-\theta)} \underbrace{\sum_{x=1}^{\infty} \dfrac{-1}{\log(1-\theta e^t)} \dfrac{(\theta e^t)^x}{x}}_{1=\sum f(x),\ \tau = e^t\theta)}$

$= \dfrac{\log(1-\theta e^t)}{\log(1-\theta)}$ for $0 < \theta e^t < 1$ or $e^t < \frac{1}{\theta}$

so sum exists — so log exists

or $t < -\log(\theta)$

38] p16 The Riemann Stieltjes integral

is $E\,h(Y) = \int_{-\infty}^{\infty} h(y)\,dF(y) = \begin{cases} \int_{-\infty}^{\infty} h(y) f(y)\,dy & Y \text{ contin} \\ \sum_{y \in \mathcal{Y}} h(y) f(y) & Y \text{ disc} \end{cases}$

39]* p16 If $F_X(x) = (1-\varepsilon) F_Z(x) + \varepsilon F_W(x)$ where $0 \le \varepsilon \le 1$ and F_Z and F_W are cdfs, then F_X is a cdf and

$$E\,g(X) = \int_{-\infty}^{\infty} g(x)\,dF_X(x) = \int_{-\infty}^{\infty} g(x)\,d\left[(1-\varepsilon)F_Z(x) + \varepsilon F_W(x)\right]$$

integration is linear $\rightarrow$ $= (1-\varepsilon)\int_{-\infty}^{\infty} g(x)\,dF_Z(x) + \varepsilon \int_{-\infty}^{\infty} g(x)\,dF_W(x).$

prop 1.14 c) So $E\,g(X) = (1-\varepsilon)\,E\,g(Z) + \varepsilon\,E\,g(W)$.

use pmf or pdf of Z to compute $E_Z\,g(Z) = E[g(Z)]$.

40] <u>Common E1 problem:</u> Find $E\,g(X)$ using 39].

ex] $F_X(x) = 0.9\,F_Z(x) + 0.1\,F_W(x)$ where $Z \sim \text{gamma}(\nu = 3, \lambda = 4)$ and $W \sim \text{Poisson}(10)$.

a) Find $E(X)$.

soln $E(X) = .9\,E(Z) + .1\,E(W) = .9\,\nu\lambda + .1\,\theta$

$= .9(3)(4) + .1(10) = 10.8 + 1 = \boxed{11.8}$

b) Find $E(X^2)$.

soln $= .9\,E(Z^2) + .1\,E(W^2)$

$= .9\left[V(Z) + (E(Z))^2\right] + .1\left[V(W) + [E(W)]^2\right]$

$= .9\left[\nu\lambda^2 + (\nu\lambda)^2\right] + .1\left[\theta + \theta^2\right]$

$= .9\left[3(16) + 9(16)\right] + .1\left[10 + 100\right]$

$= .9(192) + .1(110) = \boxed{183.8}$

41] Common error: in last ex,

$V(x) = E(X^2) - [E(X)]^2 = 183.8 - (11.8)^2 = 44.56$

but students take $Z \perp\!\!\!\perp W$ and compute $V(.9Z + .1W) = (.9)^2 V(Z) + (.1)^2 V(W)$

$= .81\,3(16) + .01(10) = 38.98.$

X has cdf $F_X(x) = .9 F_Z(x) + .1 F_W(x)$. X has a a mixture distribution. $X \neq$ $Y = .9Z + .1W$ which is a linear combination of random variables.

42] Know the 10 distributions of §1.7 = end of Exam 1 review

Ch 2 1] p29 often have n random variables $Y_1, \ldots, Y_n$ of interest.

ex] For a randomly chosen person, let Y_1 = height, Y_2 = weight, Y_3 = cholesterol level.

2] The joint pmf of $Y_1, \ldots, Y_n$ is $f(\underset{\sim}{y}) = f(y_1, \ldots, y_n) = P(Y_1 = y_1, \ldots, Y_n = y_n)$.

3] $f(x,y) = P(X=x, Y=y) = P(\{X=x\} \cap \{Y=y\})$ is often displayed with a table.

ex]

X \ Y	1	7	$P(X=x)$
9	1/7	3/7	4/7
8	2/7	1/7	3/7
$P(Y=y)$	3/7	4/7	

$P(X=9, Y=7) = \frac{3}{7}$ etc.

4] p29 $P((X,Y) \in A) = \sum_{(x,y) \in A} f(x,y)$

5] $\sum_{\underset{\sim}{y}} f(\underset{\sim}{y}) = 1$

6) p35 The joint cdf $F(y_1, \ldots, y_n)$
$= P(Y_1 \le y_1, \ldots, Y_n \le y_n)$.

7] * p30 The joint pdf $f(\underline{y}) = f(y_1, \ldots, y_n)$
satisfies $F(y_1, \ldots, y_n) = \int_{-\infty}^{y_n} \cdots \int_{-\infty}^{y_1} f(t_1, \ldots, t_n)\, dt_1 \cdots dt_n$.

8] * p30 The <u>support</u> of $Y_1, \ldots, Y_n$ for $f(\underline{y}|\underline{\theta})$, $\underline{\theta} \in \Theta$
is $\mathcal{Y}_{\underline{\theta}} = \{\underline{y} \mid f(\underline{y}|\underline{\theta}) > 0\}$. Use $\mathcal{Y}$ if the
support does not depend on $\underline{\theta}$.

9] p30 If $Y_1, \ldots, Y_n$ have joint cdf F and
joint pdf f, then

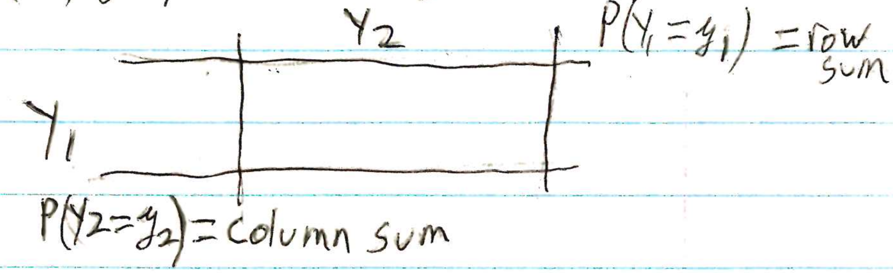
$$f(y_1, \ldots, y_n) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} F(y_1, \ldots, y_n)$$

wherever the partial derivative exists.

10] marginal: integrate or sum out coordinates not in the marginal.

11] P 30 <u>know</u> The marginal pmf
$f_{Y_1}(y_1) = \sum_{y_2} f(y_1, y_2)$ where y_1 is held fixed,

$f_{Y_2}(y_2) = \sum_{y_1} f(y_1, y_2)$ where y_2 is held fixed



12] P31 <u>know</u> The marginal pdf
$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2$ where y_1 is held fixed
$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1$ where y_2 is held fixed.

13]* p31 Let (Y_1, Y_2) have joint pmf $f(y_1, y_2)$. 6.5

The conditional pmf of Y_1 given $Y_2 = y_2$ is a function of y_1 and $f_{Y_1|Y_2=y_2}(y_1|y_2) = \dfrac{f(y_1, y_2)}{f_{Y_2}(y_2)}$, $f_{Y_2}(y_2) > 0$.

Similarly $f_{Y_2|Y_1=y_1}(y_2|y_1) = \dfrac{f(y_1, y_2)}{f_{Y_1}(y_1)}$ for $f_{Y_1}(y_1) > 0$.

14]* p32 Let (Y_1, Y_2) have joint pdf $f(y_1, y_2)$. Then the conditional pdf $f_{Y_1|Y_2=y_2}(y_1|y_2) = \dfrac{f(y_1, y_2)}{f_{Y_2}(y_2)}$ for $f_{Y_2}(y_2) > 0$, and $f_{Y_2|Y_1=y_1}(y_2|y_1) = \dfrac{f(y_1, y_2)}{f_{Y_1}(y_1)}$ for $f_{Y_1}(y_1) > 0$.

15]* A conditional pmf is a pmf, so $\sum_{y_1} f(y_1|y_2) = 1$.

A conditional pdf is a pdf, so $\int_{-\infty}^{\infty} f(y_1|y_2)\,dy_1 = 1$.

§2.2 16]* p34 $Y_1, \ldots, Y_n$ are independent if $F(y_1, \ldots, y_n) = \prod_{i=1}^{n} F_{Y_i}(y_i)$ $\forall y_i$

or if $f(y_1, \ldots, y_n) = \prod_{i=1}^{n} f_{Y_i}(y_i)$ $\forall y_i$ where f is a pmf or pdf, otherwise $Y_1, \ldots, Y_n$ are dependent.

17]* p34 The support of $\underline{Y} = (Y_1, \ldots, Y_n)$ is $\mathcal{Y} = \{\underline{y} \mid f(\underline{y}) > 0\}$. The support is a cross product if $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_n = \{\underline{y} \mid y_i \in \mathcal{Y}_i, i = 1, \ldots, n\}$ where $\mathcal{Y}_i$ is the support of Y_i.