Most engineering problems deal with multiple numerical quantitities.
Random vectors (or vector random variables) are a general tool for analyzing such random phenomena.
The development of random vectors completely parallels that of random variables:

- 4.1, 4.5: Definition
- 4.2, 4.5: Joint cdf, pdf
- 4.3, 4.5: Independence (new)
- 4.4, 4.5: Conditional cdf, pdf, Bayes rule, total probability
- 4.6: Transformations: $Z = g(X, Y)$
- 4.7: Expectation, Moments
- 4.7: Correlation, covariance (new)
- 4.8: Gaussian random vectors
- skip 4.9, 4.10

4.1

Random Vectors (or vector random variables)

Given a sample space S, a random vector $\underline{X} = (X_1, \ldots, X_n)$ is a n-tuple of random variables: $X_i : S \to \mathbb{R},\ i = 1, \ldots, n$.
Thus each sample point is mapped into a vector of n real numbers, or $\underline{X} : S \to \mathbb{R}^n$

The behaviour of a random vector is completely describes by

- the **joint probability mass function (joint pmf)** for discrete random vectors, and
- the **joint probability density function (joint pdf)** for continuous (or mixed) random vectors.

Event Shorthand

Example: $[X = 5,\ Y < 3] = \{s \in S : X(s) = 5 \text{ and } Y(s) < 3\}$, i.e., the "," denotes "and" or intersection:
$[X = 5,\ Y < 3] = [X = 5] \cap [Y < 3]$

Joint Probability Mass Function (Joint PMF) for discrete r.v.

Example: roll fair 6-sided die.

roll	1	2	3	4	5	6	(outcomes or sample points $s \in S$)
X	0	0	3	0	0	3	3 if number of dots is a multiple of 3, and 0 otherwise
Y	0	2	0	2	0	2	2 if number of dots is even, and 0 otherwise

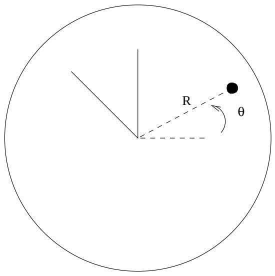
$S_X = \{0, 3\}$, $S_Y = \{0, 2\}$
PMF Shorthand: $p(x_i, y_j) = P[X = x_i, Y = y_j]$
$p(0, 0) = P(\text{ roll 1 or 5 }) = 2/6$
$p(3, 0) = P(\text{ roll 3 }) = 1/6$
$p(0, 2) = P(\text{ roll 2 or 4 }) = 2/6$
$p(3, 2) = P(\text{ roll 6 }) = 1/6$
Note that $\sum_{x_i, y_j} p(x_i, y_j) = 1$
Can answer any question, e.g. $P[XY > 5] = \sum_{\{(x,y): x \in S_X, y \in S_Y, xy > 5\}} P[X = x, Y = y] = P[X = 3, Y = 2] = 1/6$

For a random vector with two components (a pair of random variables), the joint pmf can be displayed as a **2D Histogram**

Equivalent Event approach to calculating probabilities for random vectors

Example: uniformly-likely circular dart board
Sample space: $S = \{(a,b) : \; a^2 + b^2 \le 30^2\}$
Two random variables defined on S:
- Distance from center: if $s = (a,b)$ then $R(s) = \sqrt{a^2+b^2}$
- Angular position $\Theta(s) = \tan^{-1}(b/a)$

Random vector: $\underline{X} = (R, \Theta)$ i.e. $\underline{X}(s) = (R(s), \Theta(s))$
Induced sample space: $S_{\underline{X}} = [0,30] \times [0,360]$
Questions: what is $P[R\cos\Theta \le 10]$ and $P([R \le 20] \cap [45 \le \Theta \le 90])$?
Answer: find **equivalent event** (set of sample points that satisfy the conditions)



For 2nd question, area(wedge)/area(board) $= \pi 20^2/8/(\pi 30^2) = 1/18$

General formula:

$$[\underline{X} \in B] = \{s \in S : \underline{X}(s) \in B\} = \underline{X}^{-1}(B) \quad \text{so} \quad P[\underline{X} \in B] = P(A), \text{ where } A = \{s \in S : \underline{X}(s) \in B\}$$

The latter question is said to be in "product form"
In general an event of the form $A = [X_1 \in B_1] \cap [X_2 \in B_2] \cap \cdots \cap [X_n \in B_n]$ is in **product form**
Since $[X_1 \in B_1] \cap [X_2 \in B_2] \cap \cdots \cap [X_n \in B_n] = [\underline{X} \in B]$ where $B = B_1 \times B_2 \times \cdots \times B_n$

Any event of interest can be formed from (limits of) unions of events in product form. Therefore, if we know all the probabilities of the form

$$P([X_1 \le x_1] \cap [X_2 \le x_2] \cap \cdots \cap [X_n \le x_n]) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n]$$

we can compute any probablity of interest
(although really we will use the above to get joint pdf and then integrate to get probabilities)

4.2

Joint Cumulative Probability Distribution Function (Joint CDF) (Joint Distribution)

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n] \text{ for } -\infty < x_i < \infty,\ i = 1,\ldots,n$$

Read: the probability of the event that the r.v. X_1 takes on a value in the interval $(-\infty, x_1]$ and the r.v. X_2 takes on a value in the interval $(-\infty, x_2]$ and ...

Properties of Joint CDF (for 2-vector (X,Y) only, generalization to n-vector is straightforward)

- $0 \le F_{X,Y}(x,y) \le 1,\ \forall x, y$
- $F_{X,Y}(\infty,\infty) = 1$
- $F_{X,Y}(-\infty, y) = F_{X,Y}(x,-\infty) = 0,\ \forall x, y$
- Monotone nondecreasing: $F_{X,Y}(x_1,y_1) \le F_{X,Y}(x_2,y_2)$ if $x_1 \le x_2$ and $y_1 \le y_2$
- $F_{X,Y}(x)$ is "right" continuous: $F_{X,Y}(x,y) = \lim_{\delta\searrow 0} F_{X,Y}(x+\delta,y) = \lim_{\delta\searrow 0} F_{X,Y}(x,y+\delta)$ (for $\delta > 0$)
- $P[x_1 < X \le x_2, Y \le y] = F_{X,Y}(x_2,y) - F_{X,Y}(x_1,y)$ (sketch)
- $P[X \le x,\ y_1 < Y \le y_2] = F_{X,Y}(x,y_2) - F_{X,Y}(x,y_1)$
- $P[x_1 < X \le x_2,\ y_1 < Y \le y_2] = F_{X,Y}(x_2,y_2) - F_{X,Y}(x_2,y_1) - F_{X,Y}(x_1,y_2) + F_{X,Y}(x_1,y_1)$
- $P[X = x, Y = y] = \lim_{\delta_x \to 0} \lim_{\delta_y \to 0} P[x - \delta_x < X \le x, y - \delta_y < Y \le y] = \ldots ?$

Example: circular dart board with R and Θ as defined above

$$F_{R,\Theta}(r,\theta) = P([R \le r,\ \Theta \le \theta]) = P([R \le r] \cap [\Theta \le \theta]) = \begin{cases} 0, & r < 0 \text{ or } \theta < 0 \\ (r/30)^2\theta/360, & 0 \le r \le 30, 0 \le \theta \le 360 \\ \frac{\theta}{360}, & r > 30, 0 \le \theta \le 360 \\ (r/30)^2, & 0 \le r \le 30, \theta > 360 \\ 1, & r > 30, \theta > 360 \end{cases}$$

Marginal CDF or Marginal Distribution

$F_X(x) = F_{X,Y}(x,\infty)$, $F_Y(y) = F_{X,Y}(y,\infty)$

$$F_{X,Y}(x,\infty) = P([X \le x] \cap [Y \le \infty]) = P([X \le x] \cap S) = P([X \le x]) = F_X(x)$$

In preceding example

$$F_\Theta(\theta) = F_{R,\Theta}(\infty,\theta) = \begin{cases} 0, & \theta < 0 \\ \frac{\theta}{360}, & 0 \le \theta \le 360 \\ 1, & \theta > 360 \end{cases}$$

which is Uniform(0,360): makes sense.

Proof of relation between probability and CDF

$$[x_1 < X \le x_2,\ y_1 < Y \le y_2] = ([x_1 < X \le x_2] \cap [Y \le y_2]) \cap [Y \le y_1]^c$$

since $P(A \cap B^c) = P(A) - P(A \cap B)$:

$$P[x_1 < X \le x_2,\ y_1 < Y \le y_2] = P[x_1 < X \le x_2,\ Y \le y_2] - P[x_1 < X \le x_2,\ Y \le y_1]$$

Applying same approach to X yields:

$$P[x_1 < X \le x_2,\ y_1 < Y \le y_2] = P[X \le x_2,\ Y \le y_2] - P[X \le x_1,\ Y \le y_2] - (P[X \le x_2,\ Y \le y_1] - P[X \le x_1,\ Y \le y_1])$$

$$= F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$

Discrete random vector

CDF:

$$F_{X,Y}(x, y) = \sum_{i,j} P[X = x_i, Y = y_j] u(x - x_i) u(y - y_j)$$

since

$$P[X \le x, Y \le y] = \sum_{\{i: x_i \le x\}} \sum_{\{j: y_j \le y\}} P[X = x_i, Y = y_j]$$

pdf:

$$f_{X,Y}(x, y) = \sum_{i,j} P[X = x_i, Y = y_j] \delta(x - x_i, y - y_j)$$

Computing probabilities via $P[x_1 < X \le x_2,\ y_1 < Y \le y_2] = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1)$ is inconvenient. Integrating a pdf would be easier

Joint Probability Density Function (pdf)
The joint probability density function (pdf) of a n-dimensional random vector $\underline{X}$ is defined to be

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{d^n}{dx_1 dx_2 \cdots dx_n} F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

(details about nondifferentiability similar to 1D case)

Example: Joint pdf of R, Θ for circular dartboard:

$$f_{R,\Theta}(r, \theta) = \begin{cases} \frac{2r}{30^2}\frac{1}{360}, & 0 < r < 30, 0 < \theta < 360 \\ 0, & \text{otherwise} \end{cases}$$

Properties of joint pdf (for 2-vector only) (generalization strightforward)

- $f_{X,Y}(x, y) \ge 0\ \forall x$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx\ dy = 1$
- $F_{X,Y}(x, y) = \int_{-\infty}^{x^+} \int_{-\infty}^{y^+} f_{X,Y}(x', y')\ dx'\ dy'$
- $F_X(x) = \int_{-\infty}^{x^+} \int_{-\infty}^{\infty} f_{X,Y}(x', y)\ dx'\ dy$ since $F_X(x) = F_{X,Y}(x, \infty)$
  $F_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y^+} f_{X,Y}(x, y')\ dx\ dy'$
- $P[x_1 < X \le x_2,\ y_1 < Y \le y_2] = \int_{x_1^+}^{x_2^+} \int_{y_1^+}^{y_2^+} f_{X,Y}(x, y)\ dx\ dy$
- $P[(X, Y) \in B] = \iint_B f_{X,Y}(x, y)\ dx\ dy$

If 1st two hold, then $f_{X,Y}$ is valid joint pdf

Interpretation of joint pdf
$P[x - \delta_x < X \le x,\ y - \delta_y < Y \le y] \approx \delta_x \delta_y f_{X,Y}(x, y)$ for small δ's
Higher density = more likely

Marginal pdf

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^{x^+} \int_{-\infty}^{\infty} f_{X,Y}(x', y)\ dx'\ dy = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy$$

Similarly

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx \quad \text{and in general } f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\ dx_1 \cdots dx_{i-1}\ dx_{i+1} \cdots dx_n$$

Example: Joint pdf of X, Y for circular dartboard:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi r^2}, & x^2 + y^2 \le r^2 \\ 0, & \text{otherwise} \end{cases} \quad \text{uniform over entire dartboard}$$

so $P[(X, Y) \in A] = \text{Area}(A)/(\pi r^2)$ $S_X = [-r, r]$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy = \int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} \frac{1}{\pi r^2}\ dy = \frac{2}{\pi r^2}\sqrt{r^2 - x^2} \text{ for } |x| \le r$$

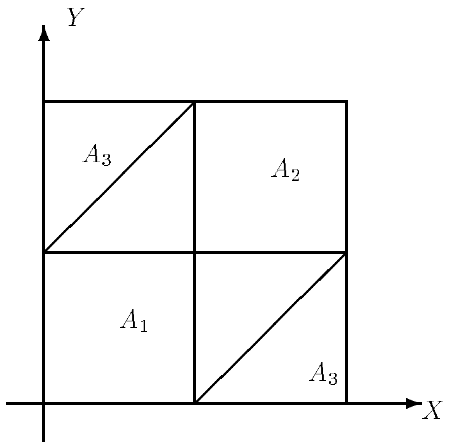
Example: a 1-meter stick breaks "randomly" into 3 pieces.
What is probability a triangle cannot be formed from 3 pieces?
Let X and Y denote break points. Assume

$$f_{X,Y}(x,y) = \begin{cases} 1, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{(2d uniform pdf)}$$

$A = \{$ cannot form triangle $\} = \{$ longest piece $> 1/2\} = A_1 \cup A_2 \cup A_3$
where $A_1 = [X < 1/2, Y < 1/2]$, $A_2 = [X > 1/2, Y > 1/2]$, $A_3 = [|X - Y| > 1/2]$

$$P(A_1) = \iint_{A_1} f_{X,Y}(x,y)\ dx\ dy = \text{Area}(A_1) = 1/4$$

$P(A) = 3/4$



Recall that events A and B are independent iff $P(A \cap B) = P(A)P(B)$
We want an analogous concept for random variables.

4.3

Independence of Random Variables

We say $X_1, \ldots, X_n$ are independent random variables iff the following *separability* conditions hold $\forall x_1, \ldots x_n \in \mathbb{R}$:

$$\begin{aligned} P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n] &= P[X_1 \le x_1]P[X_2 \le x_2] \cdots P[X_n \le x_n] \\ F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) &= F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n) \\ f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) &= f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_n}(x_n) \end{aligned}$$

Alternate equivalent definition: $P[X_1 \in B_1, \ \ldots, X_n \in B_n] = P[X_1 \in B_1] \cdots P[X_n \in B_n]$

Example: Joint pdf of R, Θ for circular dartboard:

$$f_{R,\Theta}(r, \theta) = \begin{cases} \frac{2r}{30^2}\frac{1}{360}, & 0 < r < 30, \ 0 < \theta < 360 \\ 0, & \text{otherwise} \end{cases}$$

$$f_R(r) = \begin{cases} \frac{2r}{30^2}, & 0 < r < 30 \\ 0, & \text{otherwise} \end{cases} \qquad f_\Theta(\theta) = \begin{cases} \frac{1}{360}, & 0 < \theta < 360 \\ 0, & \text{otherwise} \end{cases}$$

So R and Θ for circular dartboard. (But X and Y are dependent!)
(In contrast, for square dartboard X and Y are indepedent, but R and Θ are dependent.)

Property of Independent R.V.s

If $X_1, \ldots, X_n$ are independent r.v.s, then functions of disjoint subsets of the X_i's are independent.
E.g., if $Y_1 = X_1 + X_2, \ Y_2 = X_7 e^{X_4}, \ Y_3 = \log X_5$ then Y_1, Y_2 and Y_3 are independent r.v.s.

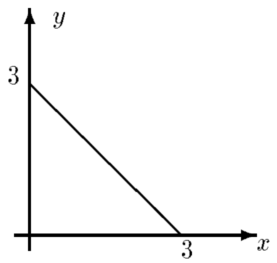
Independence arises two ways. In some problems (such as dart board above) we know how the r.v. is defined (from sample space) and we form joint CDF and then check to see if independent or not. In other problems we are given that some of the r.v.s are independent, and we use that fact to calculate other things of interest.

Example: a satellite has two power supplies, a primary and a backup that switches on if the primary fails.
The failure time X of primary (relative to launch time) is exponential with mean 1 year.
The failure time Y of backup (relative to engaging) is exponential with mean 1 year.
Assume lifetime of two supplies is independent.
What is prob. that satellite still works after 3 years?
Translate:

$$P[X + Y \ge 3] = \iint_{\{(x,y): x+y \ge 3\}} f_{X,Y}(x, y) \ dx \ dy = 1 - \iint_{\{(x,y): x+y<3\}} f_X(x) f_Y(y) \ dx \ dy$$

$$= 1 - \int_0^3 \int_0^{3-x} e^{-x} e^{-y} \ dx \ dy = 1 - \int_0^3 e^{-x}(1 - e^{-(3-x)}) \ dx = 1 - \int_0^3 e^{-x} - e^{-3} \ dx = 4e^{-3}$$

Two aspects of region to consider:
- where is $f_{X,Y}$ nonzero,
- what subset of that is the event of interest?



also: Buffon's needle

CDF $F_X(x) = P[X \leq x]$	$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P[X_1 \leq x_1,\ldots,X_n \leq x_n]$
pdf $f_X(x) = \frac{d}{dx} F_X(x)$	$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{d^n}{dx_1 \cdots dx_n} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$

Conditioning on Events

Conditional CDF $F_X(x \vert A) = P[X \leq x \vert A]$	$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n \vert A) = P[X_1 \leq x_1,\ldots,X_n \leq x_n \mid A]$
Conditional pdf (given an event) $f_X(x \vert A) = \frac{d}{dx} F_X(x \vert A)$	$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n \vert A) = \frac{d^n}{dx_1 \cdots dx_n} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n \mid A)$
Total Probability for CDF, if $\{A_i\}$ partition S $F_X(x) = \sum_i F_X(x \vert A_i) P(A_i)$	$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \sum_i F_{X_1,\ldots,X_n}(x_1,\ldots,x_n \mid A_i) P(A_i)$
Total Probability for pdf, if $\{A_i\}$ partition S $f_X(x) = \sum_i f_X(x \vert A_i) P(A_i)$	$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \sum_i f_{X_1,\ldots,X_n}(x_1,\ldots,x_n \mid A_i) P(A_i)$
Total Expectation, if $\{A_i\}$ partition S $E[g(X)] = \sum_i E[g(X) \vert A_i] P(A_i)$	$E[g(X_1,\ldots,X_n)] = \sum_i E[g(X_1,\ldots,X_n) \vert A_i] P(A_i)$

Point Conditioning and Events

Bayes rule for point conditioning $P(A \vert X = x) = f_X(x \vert A) P(A) / f_X(x)$	$P(A \vert X_1 = x_1,\ldots,X_n = x_n) = \dfrac{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n \vert A) P(A)}{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}$
Total Probability for event from pdf $P(A) = \int_{-\infty}^{\infty} P(A \vert X = x) f_X(x) \, dx$	$P(A) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(A \vert X_1 = x_1,\ldots,X_n = x_n) f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \, dx_1 \cdots dx_n$

Two random variables (NEW)

Bayes rule for two r.v.s	$f_{Y \vert X}(y \vert x) = f_{X \vert Y}(x \vert y) f_Y(y) / f_X(x)$
Total probability for two r.v.s	$f_Y(y) = \int_{-\infty}^{\infty} f_{Y \vert X}(y \vert x) f_X(x) \, dx = \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dx$
Law of iterated expectation for two r.v.s	$E[Y] = E[E[Y \vert X]] = \int_{-\infty}^{\infty} E[Y \vert X = x] f_X(x) \, dx$

4.4

Interval conditioning

$$F_Y(y \mid x_1 < X \le x_2) = P[Y \le y \mid x_1 < X \le x_2] = \frac{P[Y \le y,\ x_1 < X \le x_2]}{P[x_1 < X \le x_2]}$$

$$= \frac{\int_{-\infty}^{y} \int_{x_1}^{x_2} f_{X,Y}(x,y')\ dx\ dy'}{\int_{-\infty}^{\infty} \int_{x_1}^{x_2} f_{X,Y}(x,y')\ dx\ dy'} = \frac{\int_{-\infty}^{y} \int_{x_1}^{x_2} f_{X,Y}(x,y')\ dx\ dy'}{\int_{x_1}^{x_2} f_X(x)\ dx}$$

Define the conditional pdf

$$f_Y(y \mid x_1 < X \le x_2) = \frac{d}{dy} F_Y(y \mid x_1 < X \le x_2) = \frac{d}{dy} \frac{\int_{-\infty}^{y} \int_{x_1}^{x_2} f_{X,Y}(x,y')\ dx\ dy'}{\int_{x_1}^{x_2} f_X(x)\ dx} = \frac{\int_{x_1}^{x_2} f_{X,Y}(x,y)\ dx}{\int_{x_1}^{x_2} f_X(x)\ dx}$$

Note that naturally as $x_1 \to -\infty$ and $x_2 \to \infty$, $f_Y(y \mid x_1 < X \le x_2) \to f_Y(y)$.

Point conditioning and conditional pdfs

Recall that earlier we defined point-conditioning probabilities by

$$P(A|X = x) = \lim_{\delta \to 0} P(A \mid x - \delta < X \le x)$$

Now let A be the event $[Y \le y]$, then

$$F_{Y|X}(y|x) = P[Y \le y \mid X = x] = \lim_{\delta \to 0} P[Y \le y \mid x - \delta < X \le x] = \lim_{\delta \to 0} \frac{P[Y \le y, x - \delta < X \le x]}{P[x - \delta < X \le x]}$$

$$= \lim_{\delta \to 0} \frac{\frac{1}{\delta} \int_{-\infty}^{y} \int_{x-\delta}^{x} f_{X,Y}(x',y')\ dx'\ dy'}{\frac{1}{\delta} \int_{-\infty}^{\infty} \int_{x-\delta}^{x} f_{X,Y}(x',y')\ dx'\ dy'} = \frac{\int_{-\infty}^{y} f_{X,Y}(x,y')\ dy'}{\int_{-\infty}^{\infty} f_{X,Y}(x,y')\ dy'} = \frac{\int_{-\infty}^{y} f_{X,Y}(x,y')\ dy'}{f_X(x)}$$

Define the conditional pdf

$$f_{Y|X}(y|x) = \frac{d}{dy} F_{Y|X}(y|x) = \frac{d}{dy} \frac{\int_{-\infty}^{y} f_{X,Y}(x,y')\ dy'}{f_X(x)} = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Similarly

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

So combining yields **Bayes rule** for pdfs:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$

"Total probability" for pdfs:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\ dx$$

Calculating probability from condition pdf:

$$P[Y \in B | X = x] = \int_B f_{Y|X}(y|x)\ dy$$

Conditional pdfs in general

$$f_{X_1,X_2|X_3,X_4,X_5}(x_1,x_2 \mid x_3,x_4,x_5) = \frac{f_{X_1,X_2,X_3,X_4,X_5}(x_1,x_2,x_3,x_4,x_5)}{f_{X_3,X_4,X_5}(x_3,x_4,x_5)}$$

etc.

Chain rule for pdfs

$$f(x_1,\ldots,x_n) = f(x_n|x_1,\ldots,x_{n-1})f(x_{n-1}|x_1,\ldots,x_{n-2})\cdots f(x_2|x_1)f(x_1)$$

(with lazy notation omitting subscripts)

Independence and conditional pdfs
If X and Y are independent r.v.s, then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ so

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y) \qquad \text{equivalently} \qquad F_Y(y|X \in B) = F_Y(y)\ \forall B \in \mathcal{B}$$

Similarly

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

this gives us two alternate conditions for testing independence

---

Independence and expectation
If X and Y are independent r.v.s, then

$$E[Y|X] = E[Y] \quad \text{and} \quad E[X|Y] = E[X]$$

However, the converse is *not* true
Example: circular dart board with

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi r^2}, & x^2+y^2 \le r^2 \\ 0, & \text{otherwise} \end{cases}$$

then deriving the conditional pdf $f_{Y|X}(y|x)$ is symmetric about 0 so $E[Y|X=x]=0$. Also the marginal pdf $f_Y(y)$ is symmetric about 0 so $E[Y]=0$. Thus $E[Y|X]=E[Y]=0$. But X and Y are not independent.

Example (4.23 reworded)
Packets arriving at a router.
The number that arrive during any time interval $[t_1, t_2]$ is a Poisson r.v. with mean $\beta(t_2 - t_1)$
Assume that the amount of time needed to process the packet is a r.v. with exponential distribution with mean $1/\lambda$.
Find the PMF of N, the number of (new) packets that arrive while a specific packet is being processed.
(Implication, if N is large with high prob., then packet loss)
N is discrete r.v. with $S_N = \{0, 1, 2, \ldots\}$
Let T be the time required to process specified packet. We are given:

$$f_T(t) = \lambda e^{-\lambda t} u(t).$$

By "total prob" (eqn 4.35)

$$P[N = k] = \int_0^\infty P[N = k | T = t] f_T(t)\ dt = \int_0^\infty \frac{(\beta t)^k}{k!} e^{-\beta t} \lambda e^{-\lambda t}\ dt$$
$$= \frac{\beta^k \lambda}{k!} \int_0^\infty t^k e^{-(\beta+\lambda)t}\ dt = \frac{\beta^k \lambda}{(\beta+\lambda)^{k+1}} = \left(\frac{\lambda}{\beta+\lambda}\right)\left(\frac{\beta}{\beta+\lambda}\right)^k$$

since $\int_0^\infty x^k e^{-ax}\ dx = k!/a^{k+1}$, so

$$P[N \geq k] = \sum_{j=k}^\infty \frac{\lambda}{\beta+\lambda}\left(\frac{\beta}{\beta+\lambda}\right)^j = \left(\frac{\beta}{\beta+\lambda}\right)^k$$

Note: large β means overflow, unless λ large enough to keep up.
Specific application: suppose $\beta = 2/\mu$sec (on average 2 new packets per microsecond)
How large must λ be to ensure that only 5% of the time will 9 or more new packets arrive while during processing?
Want $P[N \geq 9] \leq 0.05 = p$
$(\frac{\beta}{\beta+\lambda})^k \leq p$ so $\lambda \geq \beta(1/\sqrt[k]{p} - 1) = 2(1/\sqrt[9]{0.05} - 1) \approx 0.79$
Need to be able to process average of $1/0.79 = 1.27$ packets per μsec

---

Conditional Expectation

$$E[Y|X = x] = \int_{-\infty}^\infty y f_{Y|X}(y|x)\ dy$$

Note that $E[Y|X = x]$ is a function that maps any value x into a number.
Call this function $g(x) = E[Y|X = x]$ Then $E[Y|X] = g(X)$ is a random variable

---

Law of Iterated Expectation

$$E[E[Y|X]] = E[Y]$$
$$E[E[Y|X]] = \int_{-\infty}^\infty E[Y|X = x] f_X(x)\ dx = \int_{-\infty}^\infty \int_{-\infty}^\infty y f_{Y|X}(y|x)\ dy f_X(x)\ dx$$
$$= \int_{-\infty}^\infty \int_{-\infty}^\infty y f_{X,Y}(x, y)\ dx\ dy = \int_{-\infty}^\infty y f_Y(y)\ dy = E[Y]$$

Example: mean of packets arriving (earlier)
Intuition: $E[N] = \beta E[T] = \beta/\lambda$

$$E[N] = \int_0^\infty E[N|T = t] f_T(t)\ dt = \int_0^\infty (\beta t) f_T(t)\ dt = \beta E[T] = \beta/\lambda$$

4.6

Functions or transformations of a random vectors
Let $X_1, \ldots, X_n$ be r.v.s whose joint pdf (perhaps via jcdf) is known.
Let $Z = g(X_1, \ldots, X_n)$ where $g : \mathbb{R}^n \to \mathbb{R}$. Want to find pdf (via cdf) of Z in terms of joint pdf of X_i's.

Method of Equivalent Events
Let

$$R_z = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : g(x_1, \ldots, x_n) \leq z\} = g^{-1}((-\infty, z])$$

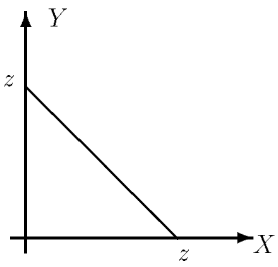
then $[Z \leq z]$ is equivalent to the event $[\underline{X} \in R_z]$, so

$$F_Z(z) = P[Z \leq z] = P[\underline{X} \in R_z] = \int \int \cdots \int_{R_z} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \;\cdots\; dx_n$$

Not really any simple plug-and-chug formula! (except Jacobian)

Typical problems for pairs of r.v.s: $Z = g(X, Y)$.
$Z = X + Y$, $Z = X \cdot Y$, $Z = X/Y$, $Z = \max(X, Y)$, $Z = \min(X, Y)$, $Z = \sqrt{X^2 + Y^2}$, $Z = |X - Y|$, $X = R \cos \Theta$, ...

Example: $Z = X + Y$
Equivalent event is easy in this case: $[Z \leq z] = [X + Y \leq z] = [Y \leq z - X] = [(X, Y) \in R_z]$ where $R_z = \{(x, y) \in \mathbb{R}^2 : x + y \leq z\}$



$$F_Z(z) = P[Z \leq z] = P[\underline{X} \in R_z] = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y)\, dy\, dx$$

gives cdf of Z in terms of joint pdf of X and Y

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx$$

gives pdf of Z in terms of joint pdf of X and Y

Sums of Independent r.v.s If X and Y are independent r.v.s, then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx = (f_X \star f_Y)(z)$$

which is a convolution integral. More generally, if $X_1, \ldots, X_n$ are independent r.v.s and $Z = X_1 + \cdots + X_n$, then

$$f_Z(z) = (f_{X_1} \star f_{X_2} \star \cdots \star f_{X_n})(z)$$

(well defined because convolution is a commutative and associative operator)

Convolution

$$f(x) = (g \star h)(x) = (h \star g)(x) = \int_{-\infty}^{\infty} g(t)h(x-t)\ dt = \int_{-\infty}^{\infty} h(t)g(x-t)\ dt$$

Example: sum of 2 independent exponential r.v.s is Erlang
Assume X and Y are independent and have exponential distributions with same mean μ
$f_X(x) = e^{-x/\mu}/\mu u(x)$
Let $Z = X + Y$ and find pdf of Z
Since range of X and Y is nonnegative, $S_Z = [0, \infty)$ so only need to consider $z \geq 0$

$$f_Z(z) = (f_X \star f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\ dx = \int_0^z \frac{1}{\mu} e^{-x/\mu} \frac{1}{\mu} e^{-(z-x)/\mu}\ dx$$

$$= \int_0^z \frac{1}{\mu^2} e^{-z/\mu}\ dx = \frac{z}{\mu^2} e^{-z/\mu} u(z)$$

so Z has an Erlang distribution.

Sums of independent r.v.s (can be shown using above convolution approach)

- $N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) \equiv N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- Poisson$\{\lambda_1\}$ + Poisson$\{\lambda_2\}$ $\equiv$ Poisson$\{\lambda_1 + \lambda_2\}$
- Binomial(n_1, p) + Binomial(n_2, p) $\equiv$ Binomial$(n_1 + n_2, p)$

Example: method of equivalent events
$Y = \max(X_1, X_2, X_3)$ where X's are assumed independent. Find pdf of Y

$$F_Y(y) = P[Y \leq y] = P[\max(X_1, X_2, X_3) \leq y] = P[X_1 \leq y, X_2 \leq y, X_3 \leq y] = F_{X_1}(y) F_{X_2}(y) F_{X_3}(y)$$

or in general:

$$= \int_{-\infty}^{y} \int_{-\infty}^{y} \int_{-\infty}^{y} f_{X_1, X_2, X_3}(x_1, x_2, x_3)\ dx_1\ dx_2\ dx_3$$

Thus by chain rule (for independent X's case):

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_{X_1}(y) F_{X_2}(y) F_{X_3}(y) + F_{X_1}(y) f_{X_2}(y) F_{X_3}(y) + F_{X_1}(y) F_{X_2}(y) f_{X_3}(y)$$

For this case the equivalent event could be found without picture. Usually a picture is needed.