36-705

- week 1, L1, L2
- Review Ch 1-3; with a view to completing problems

Ch 2 - Random variables

- sample spaces and events linked to data via random variables

2.1 - Definition

- A random variable is a mapping $X : \Omega \longrightarrow \mathbb{R}$ that assigns a real number $X(\omega)$ to each outcome ω
- Sample space is lurking in background.
- Some formalism (which I don't fully understand) ?
- Gi[illegible] a random variable X and a subset A of the real line $\mathbb{R}$
- Define $X^{-1}(A) = \{\omega \in \Omega : X(\omega) = A\}$
- And let $P(X \in A) = P(X^{-1}(A)) = P(\{\omega \in \Omega ; X(\omega) \in A\})$

$$P(X = x) = P(X^{-1}(x)) = P(\{\omega \in \Omega ; X(\omega) = x\})$$

- X denotes the random variable ; x denotes the realisation (value)

2.4 - Example - use Binomial distr.

2.2 Distribution Functions and Probability Functions

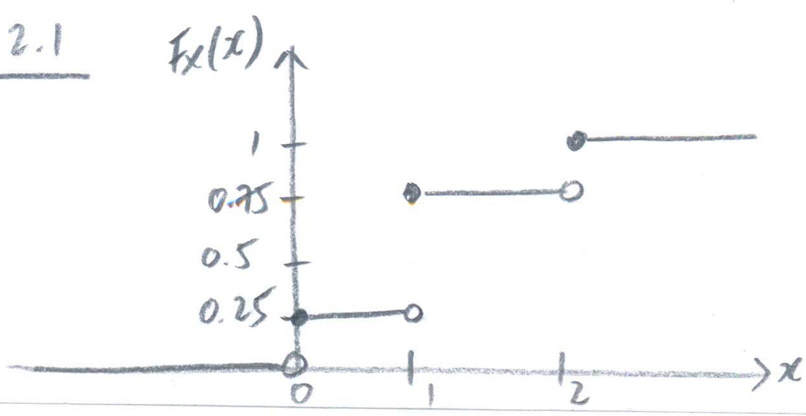
- Given a random variable X, we define the cumulative distribution function (CDF) :-

2.5 Definition (CDF)

- The CDF is the function :-

$$F_X : \mathbb{R} \longrightarrow [0,1] \quad \text{defined by} \quad F_X(x) = P(X \leqslant x) \quad (2.1)$$

- Figure 2.1



(Note the nuance of open and closed • o and repre.)

2.6 Example

- Experiment: flip coin twice; count no. heads = X
- $P(X=0) = P(X=2) = 1/4$; $P(X=1) = 1/2$; the CDF: ?

$$F_X(x) = \begin{cases} 0 & x<0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

- Note the upper bounds are closed

- Right continuous, non-decreasing, defined for all $x \in \mathbb{R}$, even though r.v. takes values 0, 1, 2
- $F_X(1.4) = 0.75 \longrightarrow P(X<1.4) = P(X=0) + P(X=1) = \frac{1}{4} + \frac{1}{2} = 0.75$
- The following results $\longrightarrow$ CDF completely determines distn of r.v.

2.7 Theorem

- Let X have CDF F and Y have CDF G. If $F(x) = G(x)$ for all x then $P(X \in A) = P(Y \in A)$ for all A.

2.8 Theorem

- A function F mapping the real line $\mathbb{R}$ to $[0,1]$ is a CDF for some probability P if and only if F satisfies the following 3 conditions:-
  - i) F is non-decreasing $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
  - ii) F is normalised $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
  - iii) F is right-continuous: $F(x) = F(x^+)$ for all x, where ?

$$F(x^+) = \lim_{\substack{y \to x \\ y > x}} F(y)$$

- unfamiliar with right-continuous:
  - internet def: $f(x)$ is continuous from the right at a iff

$$\lim_{x \to a^+} f(x) = f(a)$$

☆

Proof: - I don't fully understand the proof; will come back to this later

2.9 Definition (Discrete r.v, PMF)

- X is discrete if it takes countably many values $\{x_1, x_2, \ldots\}$
- we define a probability function or probability mass function for X by

$$f_X(x) = P(X = x)$$

- Hence $f_X(x) \geq 0$ for all $x \in \mathbb{R}$ · notation: f instead of $f_X(x)$

and $\sum_i f_X(x) = 1$

- For discrete X, the CDF and PDF are related by:-

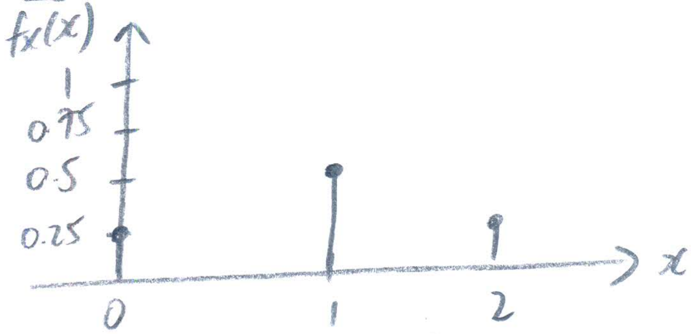
$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} f_X(x)$$

2.10 Example

- P. for example 2.6:-

$$f_X(x) = \begin{cases} 1/4 & x = 0 \\ \frac{1}{2} & x = 1 \\ 1/4 & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

figure 2.2



2.11. Definition

- A random variable X is continuous if there exists a function f_X such that:-

i) $f_X(x) \geq 0 \ \forall x$ (non-negativity)

ii) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

iii) for every $a \leq b$

$$P(a < X < b) = \int_a^b f_X(x)\, dx \quad (2.2)$$

- The function f_X is called the probability density function PDF and

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

- And $f_X(x) = F_X'(x)$ at all points x at which F_X is differentiable

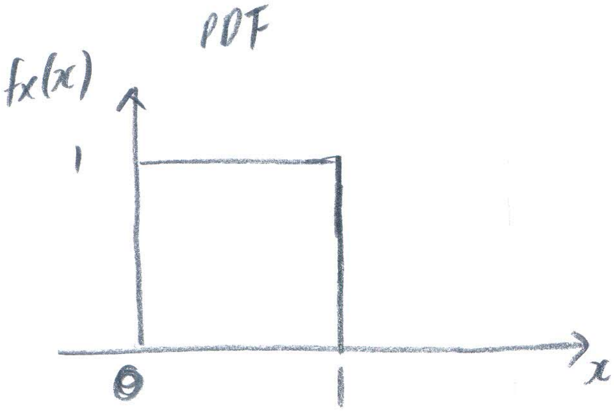
- notation $\int f(x)\, dx$, $\int f \Leftrightarrow \int_{-\infty}^{\infty} f(x)\, dx$ ⊛

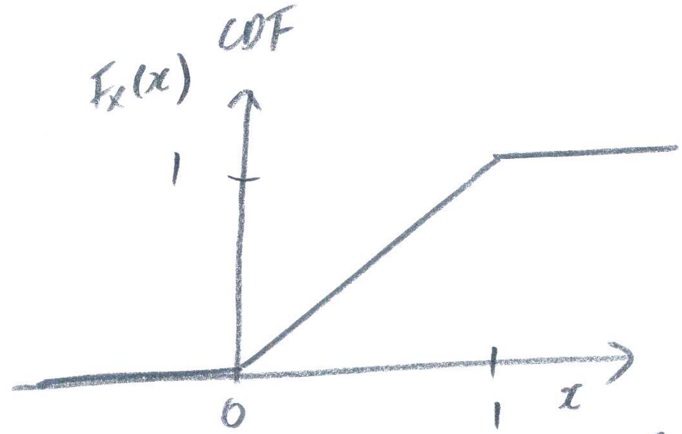
2.12 Example ? - Can't see how to integrate standard uniform PDF over $(-\infty,\infty)$ to get 1 ? (ii) ✓

2.12 Example (uniform$(0,1)$)

- X has PDF: $f_X(x) = \begin{cases} 1 & \text{for } 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$ · CDF → $F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 < x < 1 \\ 1 & x > 1 \end{cases}$

· $f_X(x) \ge 0$ ✓, $\int f_X(x)\,dx = 1$ (?) ✓





· (ii): In order to check if a PDF is well defined; see definition, in particular (ii)
- use wolfram to assist
- (x): remember for continuous X, $P(X=x) = 0 \ \forall x$ (infinitesimal point)
- Don't think of $P(X=x)$ as $f_X(x)$ for continuous, only discrete
- probabilities from integrating
- A PDF evaluated at a single point can be greater than 1; unlike a PMF
- E.g. $f(x) = 5$ for $x \in [0, \frac{1}{5}]$ and 0 otherwise; $f(x) \ge 0$ and $\int f(x)\,dx = 1$, even though $f(x) = 5$ in some places

(x): PDFs can be unbounded e.g. $f(x) = \frac{2}{3}x^{-1/3}$ for $0 < x < 1$, $f(x) = 0$ otherwise

$\int f(x)\,dx = 1$ even though unbounded

2.15 Lemma

- Let F be a CDF for an r.v. X, then:-

1. $P(X=x) = F(x) - F(x^-)$ where $F(x^-) = \lim_{y \uparrow x} F(y)$
2. $P(x < X \le y) = F(y) - F(x)$
3. $P(X > x) = 1 - F(x)$
4. X continuous then $F(b) - F(a) = P(a < X < b) = P(a \le X < b)$
$= P(a < X \le b) = P(a \le X \le b)$

· Also, the inverse CDF

2.16 Definition (Inverse CDF)

- Let X be a random variable with CDF F. The inverse CDF or quantile function is defined by:-

$$F^{-1}(q) = \inf\{x : F(x) > q\} \quad \text{for } q \in [0,1]$$

- If F is strictly increasing and continuous then $F^{-1}(q)$ is the unique real no x such that $F(x) = q$

· $F^{-1}(1/4)$ - 1st quartile ; $F^{-1}(1/2)$ - median ; $F^{-1}(3/4)$ - third quartile

· 2 r.v.s. are equal in distribution $X \stackrel{d}{=} Y$ if $F_X(x) = F_Y(x)\ \forall x$

· $X \stackrel{d}{=} Y \not\Rightarrow X = Y$

· only that all probability statements about X and Y will be the same

· E.g. suppose $P(X=1) = P(X=-1) = 1/2$

· Let $Y = -X$; then $P(Y=1) = P(Y=-1) = 1/2$ and so $X \stackrel{d}{=} Y$ but $X \neq Y$

· $P(X=Y) = 0$

2.3. Important Discrete Random Variables

· Notation: $X \sim F$ means X has distribution F, not is 'approximately'

Point Mass Distribution

- X has a point mass distribution at a : $X \sim \delta_a$ if $P(X=a) = 1$ in which case

$$F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$$

- PMF : $f(x) = 1$ for $x = a$ and 0 otherwise

Discrete Uniform Distribution

· Let $k > 1$ be a given integer

- Suppose X has PMF:-

$$f(x) = \begin{cases} 1/k & x = 1, \dots, k \\ 0 & \text{otherwise} \end{cases}$$

We say X has a uniform distribution on $\{1, \dots, k\}$

Bernoulli Distribution

- let X be a binary coin flip
- Then $P(X=1)=p$ $P(X=0)=1-p$ for some $p \in [0,1]$
- We say X has a Bernoulli distri :- $X \sim Bern(p)$
- PMF :- $f(x) = p^x(1-p)^{1-x}$ $x \in \{0,1\}$

Binomial Distribution

- Suppose coin that falls heads up with probability p for some $0 \leq p \leq 1$
- flip coin n times and let X be no. heads
- Assume independent tosses; let $f(x) = P(X=x)$ be PMF :-

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

- We say that the r.v. is a Binomial r.v. $X \sim Bin(n,p)$
- Also; if $X_1 \sim Binomial(n_1, p)$, $X_2 \sim Binomial(n_2, p)$ then $X_1 + X_2 \sim Binomial(n_1+n_2, p)$

- Random variables, parameters

- X is a random variable
- x is a particular value (realisation of the r.v.)
- n and p are (in the frequentist paradigm) parameters i.e. fixed real nos
- p is unknown and must be estimated from data $\longrightarrow$ statistical inference

Geometric Distribution

- X has a geometric distribution with paramete $p \in (0,1)$: $X \sim Geom(p)$ if

$$P(X=k) = p(1-p)^{k-1} \quad k \geq 1$$

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

- A suitable model ~~for~~ instanti of this: no. flips needed until the first head when flipping a coin.

Poisson Distribution

- X has a Poisson distri with parameter λ : $X \sim Po(\lambda)$ if

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x \geq 0$$

Note that: $\sum_{x=0}^{\infty} f(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$

- Used as a model for counts of rare events e.g radioactive decay, car accidents
- If $X_1 \sim Po(\lambda_1)$ and $X_2 \sim Po(\lambda_2)$. $X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$
- On the r.v terminology
- We defined r.v.s as mappings from a sample space Ω to $\mathbb{R}$
- The sample space often gets backgrounded here, but it is 'there'.
- An example of explicit sample space for Bernoulli
- Let $\Omega = [0,1]$
- Define P to satisfy $P([a,b]) = b-a$ for $0 \leq a \leq b \leq 1$

(?) - unfamiliar and novel territory

- Fix $p \in [0,1]$ and define:-

$$X(\omega) = \begin{cases} 1 & \omega \leq p \\ 0 & \omega > p \end{cases}$$

- Then $P(X=1) = P(\omega \leq p) = P([0,p]) = p$ and $P(X=0) = 1-p$
- Thus $X \sim Bernoulli(p)$
- In practice, we think of a random variable is a random number
- Formally: It is a mapping from $\Omega \to \mathbb{R}$

2.4 Important Continuous Random Variables

- Uniform Distribution
- $X \sim Uniform(a,b)$ if:-

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases} \quad a < b \qquad F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

Normal (Gaussian) Distribution

- X has a Normal/Gaussian with parameters μ and σ^2: $X \sim N(\mu, \sigma^2)$ if

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad x \in \mathbb{R} \qquad (2.3)$$

- where $\mu \in \mathbb{R}$, $\sigma > 0$
- Intuitively parameter μ is the 'center'/mean; σ is 'spread'/standard dev.
- Both μ and σ formally defined next.
- Many phenomena in nature have approximately Normal distr
- CLT $\Rightarrow$ distribution of sum of r.v.s can be approximated by a Normal
- X has a standard normal distr if $\mu = 0, \sigma = 1$; conventionally, Z
- PDF and CDF of standard Normal: $\phi(z)$ and $\Phi(z)$
- There is no closed form expression for $\Phi(z)$, expressed as erf special function
- useful facts:

i) If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

ii) If $Z \sim N(0,1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

iii) If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ are independent then:-

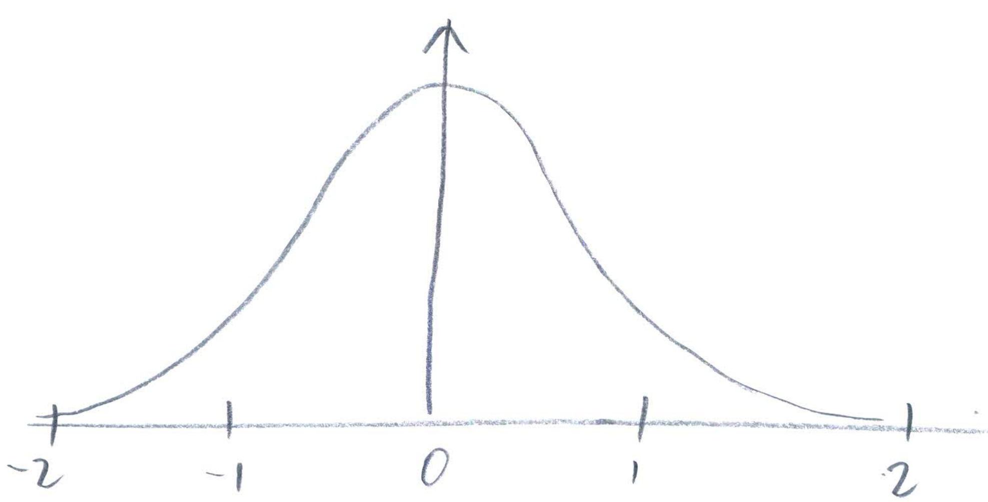
$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$$

From (i); if $X \sim N(\mu, \sigma^2)$ then

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

We can compute probabilities as long as we can compute CDF of a standard normal

All statistics packages have $\Phi(z)$ and $\Phi^{-1}(q)$



Exponential Distribution

- X has exponential distri with parameter β: $X \sim Exp(\beta)$ if:-

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \quad , x > 0 , \beta > 0$$

- Model ~~for~~ used for lifetimes of electronic components, waiting times between rare events.

Gamma Distribution

- for $\alpha > 0$, the gamma function is defined by $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} \, dy$
- X has a Gamma distri with parameters α and β, denoted $X \sim Gamma(\alpha, \beta)$ if

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad , x > 0 \quad \text{where } \alpha, \beta > 0$$

- The exponential distri is $Gamma(1, \beta)$ distri
- If $X_i \sim Gamma(\alpha_i, \beta)$ are independent then $\sum_{i=1}^n X_i \sim Gamma\left(\sum_{i=1}^n \alpha_i, \beta\right)$

Beta Distribution

- X has a Beta distri with parameters α, β, $\alpha > 0, \beta > 0$: $X \sim Beta(\alpha, \beta)$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad , 0 < x < 1$$

t and Cauchy distri

- X has a t-distri with ν degrees of freedom : $X \sim t_\nu$ if:-

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}}$$

(i): t-distri similar to normal, but has thicker tails

(ii): normal corresponds to a t-distri with $\nu = \infty$

(iii): The Cauchy distri is a special case of t-distri with $\nu = 1$

$$f(x) = \frac{1}{\pi(1+x^2)}$$

To see that this is a density:-

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{(1+x^2)}\,dx = \int_{-\infty}^{\infty} \arctan(x)\,dx$$

$$= \frac{1}{\pi}\left[\arctan x\right]_{\lim_{x\to-\infty}}^{\lim_{x\to\infty}} = \frac{1}{\pi}\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = 1$$

χ^2 distribution

- X has a χ^2 distn with p degrees of freedom : $X \sim \chi_p^2$ if:-

$$f(x) = \frac{1}{\Gamma(p/2)\,2^{p/2}}\, x^{p/2 - 1} e^{-x/2} \qquad x > 0$$

(H): If $z_1, z_2, \ldots, z_p$ are indep, standard normal r.v.s then $\sum_{i=1}^{p} z_i^2 \sim \chi_p^2$ → tomorrow

(W) (A): Plot all of these on wolfram using red-coded and built-in! or wiki

- Knowing/being familiar with shape of PDFs and CDFs will 'animate formulae'

Bivariate Distributions

- Given a pair of discrete r.v.s X and Y, define the joint prob mass function :-

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

- we write $f_{X,Y}$ when we want to be more explicit

2.18 Example

- Bivariate for 2 r.v.s.
- $f(1,1) = P(X=1, Y=1) = 4/9$

	$Y=0$	$Y=1$	
$X=0$	$1/9$	$2/9$	$1/3$
$X=1$	$2/9$	$4/9$	$2/3$
	$1/3$	$2/3$	1

2.19 Definition (Bivariate PDF)

- In the continuous case; we call a function $f(x,y)$ a PDF for r.v.s. (X,Y) if

i) $f(x,y) \geq 0 \;\forall\; (x,y)$

ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

iii) for any set $A \subset \mathbb{R}\times\mathbb{R}$, $P((X,Y)\in A) = \iint_A f(x,y)\,dx\,dy$

- In the discrete and continuous case; the joint CDF is:-

- $F_{X,Y}(x,y) = P(X \le x, Y \le y)$

- 2.6 Marginal Distributions

2.23 Definition (Marginals)

- If X and Y have joint distribution with mass function $f_{X,Y}$ then

- The marginal mass function for X:-

$$f_X(x) = P(X=x) = \sum_y P(X=x, Y=y) = \sum_y f(x,y) \quad (2.4)$$

- The marginal mass function for Y:-

$$f_Y(y) = P(Y=y) = \sum_x P(X=x, Y=y) = \sum_x f(x,y) \quad (2.5)$$

2.24 Example

- $f_{X,Y}$ specified:-

	Y=0	Y=1	
X=0	1/10	2/10	3/10
X=1	3/10	4/10	7/10
	4/10	6/10	1

(1): marginal distri for X corresponds to row totals
marginal distri for Y —— " —— to column totals

e.g. $f_X(0) = P(X=0, Y=0) + P(X=0, Y=1) = \frac{1}{10} + \frac{2}{10} = \frac{3}{10}$

$f_X(1) = P(X=1, Y=0) + P(X=1, Y=1) = \frac{3}{10} + \frac{4}{10} = \frac{7}{10}$

- $f_X(0) + f_X(1) = \frac{3}{10} + \frac{7}{10} = 1$

And $f_Y(0) = P(X=0, Y=0) + P(X=1, Y=0) = \frac{1}{10} + \frac{3}{10} = \frac{4}{10}$

$f_Y(1) = P(X=0, Y=1) + P(X=1, Y=1) = \frac{2}{10} + \frac{4}{10} = \frac{6}{10}$

$f_Y(0) + f_Y(1) = 1$

2.25 Definition (continuous marginals) - examples 2.26, 2.27, 2.28 ✓

- For continuous r.v.s; the marginal densities:-

$$f_X(x) = \int f(x,y)\,dy \qquad f_Y(y) = \int f(x,y)\,dx \quad (2.6)$$

- corresponding marginal distributions $F_X(x)$ and $F_Y(y)$

2.7 Independent Random variables

2.9 Definition (independent r.v.s)

- Two random variables X, Y are independent if:-

$$P(X\in A, Y\in B) = P(X\in A)P(Y\in B) \quad (2.7)$$

and we write $X \perp\!\!\!\perp Y$. otherwise we say that they are dependent and we write $X \not\perp\!\!\!\perp Y$

(NB): in principle, to check whether X and Y are independent, we need to check (2.7) for all subsets A and B. A useful, but not entirely rigorous result for continuous and discrete r.v.s:-

2.30 Theorem

- Let X and Y have joint PDF $f_{X,Y}$. Then $X \perp\!\!\!\perp Y$ if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all values x and y

2.31. Example

- Let X and Y have following distri:-
- $f_X(0) = f_X(1) = f_Y(0) = f_Y(1) = \frac{1}{2}$

	$Y=0$	$Y=1$	
$X=0$	$1/4$	$1/4$	$1/2$ $f_X(0)$
$X=1$	$1/4$	$1/4$	$1/2$ $f_X(1)$
	$1/2$	$1/2$	1
	$f_Y(0)$	$f_Y(1)$	

(NB): X and Y are independent:

as $f_X(0)f_Y(0) = f(0,0)$ $\quad f_X(1)f_Y(0) = f(1,0)$

$f_X(0)f_Y(1) = f(0,1)$ $\quad f_X(1)f_Y(1) = f(1,1)$

Counterexample

- $f_X(0)f_Y(0) = \left(\frac{1}{2}\right)^2 \neq f(0,0)$

	$Y=0$	$Y=1$	$f_X(x)$
$X=0$	$1/2$	0	$1/2$
$X=1$	0	$1/2$	$1/2$
$f_Y(y)$	$1/2$	$1/2$	

- Another result for verifying independence

2.33 Theorem

- Suppose that the range of X and Y is a (possibly infinite) rectangle
- If $f(x,y) = g(x)h(y)$ for some g and h (not necessarily PDFs) then X and Y are independent.

2.34 Example

- Let X and Y have density

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & x>0,\ y>0 \\ 0 & \text{otherwise} \end{cases}$$

- Range of X and Y is rectangle $(0,\infty)\times(0,\infty)$ ✓
- We can write $f(x,y) = g(x)h(y)$ where $g(x) = 2e^{-x}$ $h(y) = e^{-2y}$
- Thus $X \perp\!\!\!\perp Y$

2.8. Conditional Distributions

- X and Y discrete, then can compute conditional distribution of X given that we have observed $Y=y$.
- Specifically $P(X=x \mid Y=y) = \dfrac{P(X=x, Y=y)}{P(Y=y)}$

2.3.5 Definition (Conditional PMF)

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

if $f_Y(y) > 0$

- In the continuous case, we use same definitions *
- Note that when we compute $P(X\in A \mid Y=y)$ in the continuous case ⇒ conditioning on the event $\{Y=y\}$ with probability 0
- We avoid this problem by defining things in terms of PDF

2.36. Definition

- for continuous r.v.s ; the conditional probability density function :-

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

- Analogous, but NOT equivalent to Bayes

- assuming $f_Y(y) > 0$; then

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y)\,dx$$

(N): include more examples on this :-

2.37 Example

?

- let X and Y have a joint distribution on the unit square

i.e. $f(x,y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases} \Rightarrow f_{X|Y}(x|y) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

- Given $Y = y$, X is uniform$(0,1)$

$f_Y(y) = \int_0^1 f(x,y)\,dx = \int_0^1 1\,dx = [x]_0^1 = 1 \quad 0 \le y \le 1 \quad f_Y(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

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

2.38 Example (OT) → see how this works

from E.2.27 $\quad f_Y(y) = \begin{cases} \frac{1}{2} + y & 0 \le y \le 1 \\ & \text{otherwise} \end{cases}$

- let $f(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

- Evaluate $P(X < \frac{1}{4} | Y = \frac{1}{3})$

(W): (how about regions?)

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{x+y}{y+\frac{1}{2}}$$

$$P(X < \tfrac{1}{4} | Y = \tfrac{1}{3}) = \int_0^{1/4} f_{X|Y}(x|\tfrac{1}{3})\,dx = \int_0^{1/4} \tfrac{6}{5}(x + \tfrac{1}{3})\,dx$$

$$= \tfrac{5}{6}\left[\tfrac{1}{2}x^2 + \tfrac{1}{3}x\right]_0^{1/4} = \tfrac{6}{5}\left[\left(\tfrac{1}{32} + \tfrac{1}{12}\right)\right] = \tfrac{11}{80}$$

2.39 Example

- suppose $X \sim$ uniform$(0,1)$
- obtain a value for X, generate $Y|X = x \sim$ uniform$(0,1)$
- what is the marginal distribution of Y? ?
- note $f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ $\quad f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{if } 0<x<y<1 \\ 0 & \text{otherwise} \end{cases}$

- joint $f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x) = \begin{cases} \frac{1}{1-x} & \text{if } 0<x<y<1 \\ 0 & \text{otherwise} \end{cases}$

- marginal for Y ?

$$f_Y(y) = \int_0^y f_{X,Y}(x,y)\,dx = \int_0^y \frac{1}{1-x}\,dx$$

· set $u = 1-x$ $\quad \frac{du}{dx} = -1 \quad \frac{dx}{du} = -1$

$$= -\int_1^{1-y} \frac{1}{u}\,du = -\big[\ln u\big]_1^{1-y} = -\ln(1-y) \checkmark$$

W: Not sure how $f_{Y|X}(y|x)$ deduced

2.40 Example

- consider density in E.2.28 i.e. $f(x,y) = \begin{cases} 21/4\, x^2 y & x^2 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$
- let's find $f_{Y|X}(y|x)$ ✓
- when $X = x$, y must satisfy $x^2 \le y \le 1$ - consider bounds
- Note from E.2.28: $f_X(x) = \begin{cases} 21/8\, x^2(1-x^4) & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

W: need to master/review these examples on bounds

- invoke formula:-

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{21/4\, x^2 y}{21/8\, x^2(1-x^4)} = \frac{2y}{(1-x^4)}$$

- compute $P(Y \ge 3/4 | X = 1/2)$; noting that $f_{Y|X}(y|1/2) = 32y/15$ ✓

$$P(Y \ge 3/4 | X = 1/2) = \int_{3/4}^1 f_{Y|X}(y|1/2)\,dy = \int_{3/4}^1 32y/15 = 32/15 \left[\tfrac{1}{2}y^2\right]_{3/4}^1 = \frac{32}{15}\left(\frac{1}{2} - \frac{9}{32}\right)$$

$$= \frac{7}{15} \checkmark$$

2.9. Multivariate Distributions & IID Samples

- Let $\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ where $X_1, \ldots, X_n$ are r.v.s. ; then $\underline{X}$ is a random vector
- Let $f(x_1, \ldots, x_n)$ denote the PDF
- It is possible to define marginals, conditionals in same way as bivariate case
- We say $X_1, \ldots, X_n$ are independent if for every $A_1, \ldots, A_n$:-

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i) \quad (2.8)$$

- It suffices to check that:-

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i)$$

2.41 Definition (

- If $X_1, \ldots, X_n$ are independent and each has same marginal distr with CDF F,

we say that $X_1, \ldots, X_n$ are independent and identically distributed

- Notation:- $X_1, \ldots, X_n \sim F$
- If F has density f, we also write $X_1, \ldots, X_n \sim f$
- And $X_1, \ldots, X_n$ may be called a random sample of size n from F.

- IID is a cornerstone reference for statistical theory and practice.

2.10. Two important multivariate distr

Multinomial

- A multivariate version of the Binomial

(*) - model - consider drawing a ball from an urn which has k balls "color 1, colour 2, ..., colour k"
  - Let $p = (p_1, \ldots, p_k)$ where $p_j \geq 0$ and $\sum_{j=1}^{k} p_j = 1$
  - p_j - probability of drawing ball of colour j
  - Draw n times (independent draws with replacement)
  - Let $X = (X_1, \ldots, X_k)$ be a random vector with X_j - no. of times color j ball is drawn.
  - Hence $n = \sum_{j=1}^{k} X_j$

Then we say:

- $\underline{X}$ has a Multinomial $(n, \underline{p})$: $X \sim \text{Multinomial}(n, \underline{p})$

$$f(\underline{x}) = \binom{n}{x_1 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} \quad \text{where} \quad \binom{n}{x_1 \cdots x_k} = \frac{n!}{x_1! \ldots x_k!}$$

2.42 Lemma (Multinomial-Binomial)

- Suppose $X \sim \text{Multinomial}(n, \underline{p})$ where $\underline{X} = (X_1, \ldots, X_k)$ and $\underline{p} = (p_1, \ldots, p_k)$
- The marginal distribution of X_j is Binomial (n, p_j)

(1): So, explicitly, $f_{X_j}(x_j) = \sum_{x_1} \cdots \sum_{x_k} f(X_1 = x_1, \ldots, X_j = x_j, \ldots, X_k = x_k)$

where the summations do NOT include $\sum_{x_j}$

Multivariate Normal

- Univariate Normal has 2 parameters: μ, σ
- In the multivariate Normal, μ is a vector (rather than scalar)
  σ is a matrix (—— " ——)
- Let $\underline{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix}$ where $Z_1, \ldots, Z_k \sim N(0,1)$ are independent
- The density of $\underline{Z}$ is:- $f(\underline{z}) = \prod_{i=1}^{k} f(z_i) = \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2} \sum_{j=1}^{k} z_j^2\right\}$

$$= \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2} \underline{z}^T \underline{z}\right\}$$

- Then we say that $\underline{Z}$ has a standard Multivariate Normal: $\underline{Z} \sim N(\underline{0}, \underline{I})$
- where $\underline{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^k$ and $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$

More generally: X has a MVN distn: $X \sim N(\mu, \Sigma)$ if it has density:-

$$f(\underline{x}; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})\right\}$$

- As Σ is symmetric and positive definite i.e. $\Sigma \in \mathbb{R}^{k \times k}$ and $\Sigma \in S_k^{++}$
- It can be shown that there exists $\Sigma^{1/2}$ with following properties:-
  i) $\Sigma^{1/2}$ is symmetric

2.47 Example

- Let $X \sim$ uniform$(-1,3)$. Find the PDF of ~~x^2~~ $Y = X^2$
- The density of X is:- $f_X(x) = \begin{cases} 1/4 & \text{if } -1<x<3 \\ 0 & \text{otherwise} \end{cases}$ $\Rightarrow$ (w) $F_X(x) = \begin{cases} 0 & x<-1 \\ 1/4\, x & -1\leq x\leq 3 \\ 1 & x>3 \end{cases}$

- note Y can only take values in $(0,9)$ ✓
- Split into two cases:- (?) i) $0<y<1$ ii) $1\leq y<9$

- case (i): $A_y = \{x: x^2 \leq y\} = \{x: -\sqrt{y} \leq x \leq \sqrt{y}\}$, ~~$F_Y(y)$~~ $0<y<1$

$$F_Y(y) = P(Y\leq y) = P(X^2 \leq y) = P(\{x: -\sqrt{y}\leq x \leq \sqrt{y}\})$$

$$= \int_{\{x: -\sqrt{y}\leq x\leq \sqrt{y},\, 0<y<1\}} f_X(x)\,dx = P(-\sqrt{y}\leq x \leq \sqrt{y}) ;\ 0<y<1 \Leftrightarrow -1\leq x\leq 1$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \tfrac{1}{4}\sqrt{y} + \tfrac{1}{4}\sqrt{y} = F_Y(y) = \tfrac{1}{2}\sqrt{y} \quad 0<y<1$$

case (ii) $A_y = \{x: x^2\leq y\} = \{x: -\sqrt{y}\leq x \leq \sqrt{y}\}$, $1\leq y<9$

$$F_Y(y) = P(Y\leq y) = P(X^2\leq y) = P(\{x: \sqrt{-y}\leq x\leq \sqrt{y}\})$$

$$= \int_{A_y = \{x: \sqrt{-y}\leq x\leq\sqrt{y},\, 1\leq y<9\}} f_X(x)\,dx = P(-\sqrt{y}\leq x\leq\sqrt{y})$$

(?) (Q) - supplement-review

(x) I don't understand how we got

$A_y = [-1, \sqrt{y}]$ $F_Y(y) = \int_{A_y} f_X(x)\,dx$

$= \tfrac{1}{4}(\sqrt{y}+1)$

- When $1\leq y<9$; $-3\leq x\leq 3$;
- For $-3\leq x<-1$; $1<y\leq 9$
- For $-1\leq x\leq 3$; $1<$

(?)

- Differentiating F:

$$f_Y(y) = \begin{cases} \dfrac{1}{4\sqrt{y}} & 0<y<1 \\ \dfrac{1}{8\sqrt{y}} & 1<y<9 \\ 0 & \text{otherwise} \end{cases}$$

- When r is strictly monotone increasing or decreasing, then r has an inverse $s = r^{-1}$ and one can show:-

$$f_Y(y) = f_X(s(y)) \left| \frac{ds(y)}{dy} \right| \quad (2.12)$$

2.12. Transformations of several Random variables

- In some cases we are interested in transformations of several r.v.s.
- E.g, if X and Y are given r.v.s we may want to know distn of $X+Y$, X/Y, $\max\{X,Y\}$ or $\min\{X,Y\}$
- Define $Z = r(X,Y)$ as the function of interest

Three steps for transform:

1. For each z, find the set $A_z = \{(x,y) : r(x,y) \leq z\}$
2. Find the CDF:-

$$F_Z(z) = P(Z \leq z) = P(r(X,Y) \leq z)$$
$$= P(\{(x,y) : r(x,y) \leq z\})$$
$$= \iint_{A_z} f_{X,Y}(x,y)\, dx\, dy$$

3. Then $f_Z(z) = F_Z'(z)$

(?) - mitsque, internalise strategy

2.15 Example → see scribblings (crucial to understand)

- 2.13 Appendix

- Recall, a probability measure P is defined on a σ-field $\mathcal{A}$ of a sample space Ω
- A random variable X is a measurable map $X : \Omega \to \mathbb{R}$
- measurable → for every x $\{\omega : X(\omega) \leq x\} \in \mathcal{A}$.