ST2132
AY23/24 SEM 1
github/jovyntls

01. PROBABILITY

- **probability** of an event $\rightarrow$ the limiting relative frequency of its occurrence as the experiment is repeated many times
- the **realisation** x is a constant, and X is a generator
  - running r experiments gives us r realisations $x_1, \dots, x_r$

expectation

expectation of X

discrete: mass function	**continuous**: density function
$E(X) := \sum_{i=1}^{n} x_i p_i$	$E(X) := \int_{-\infty}^{\infty} x f(x)\,dx$

expectation of a function $h(X)$

$$E\{h(X)\} = \begin{cases} \sum_{i=1}^{n} h(x_i)p_i & X \text{ is discrete} \\ \int_{-\infty}^{\infty} h(x)f(x)\,dx & X \text{ is continuous} \end{cases}$$

variance

variance, $\mathrm{var}(X) := E\{(X-\mu)^2\}$
standard deviation, $SD(X) := \sqrt{\mathrm{var}(X)}$

law of large numbers

LLN: for a function h, as number of realisations $r \to \infty$,
$\bar{x} \to E(X)$, $v \to \mathrm{var}(X)$
$\frac{1}{r}\sum_{i=1}^{r} h(x_i) \to E\{h(X)\}$

mean of realisations, $\bar{x} := \frac{1}{r}\sum_{i=1}^{r} x_i$

variance of realisations, $v := \frac{1}{r}\sum_{i=1}^{r}(x_i - \bar{x})^2$

Monte Carlo approximation

$$E\{h(X)\} \approx \frac{1}{r}\sum_{i=1}^{r} h(x_i)$$

by LLN, as $r \to \infty$, the approximation becomes exact

joint distribution

- **discrete**: mass function
  $\Pr(X = x_i, Y = y_j) = p_{ij}$ where $x_1, \dots, x_i$ and $y_1, \dots, y_j$ are all possible values of X and Y
- **continuous**: density function
  $f : \mathbb{R}^2 \to [0, \infty), \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

for $h : \mathbb{R}^2 \to \mathbb{R}$,
$E\{h(X,Y)\} =$

$$\begin{cases} \sum_{i=1}^{I}\sum_{j=1}^{J} h(x_i, y_j)p_{ij} & X \text{ is discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)\,dx\,dy & Y \text{ is continuous} \end{cases}$$

algebra of RV's

let X, Y be RVs and a, b, c be constants
- $Z = aX + bY + c$ is also an RV
  - $z = ax + by + c$ is a realisation of Z
- linearity of expectation - $E(Z) = aE(X) + bE(Y) + c$

covariance

let $\mu_X = E(X), \mu_Y = E(Y)$.

covariance, $\mathrm{cov}(X,Y) = E\{(X-\mu_X)(Y-\mu_Y)\}$

- $\mathrm{cov}(X,Y) = E(XY) - \mu_X\mu_Y$
- $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$
- $\mathrm{cov}(X,X) = \mathrm{var}(X)$
- $\mathrm{cov}(W, aX + bY + c) = a\,\mathrm{cov}(W,X) + b\,\mathrm{cov}(W,Y)$
- $\mathrm{var}(aX + bY + c) =$
  $a^2\,\mathrm{var}(X) + b^2\,\mathrm{var}(Y) + 2ab\,\mathrm{cov}(X,Y)$

joint, marginal & conditional distributions

let $f(x,y)$ be the **joint** density and $f_X(x), f_Y(y)$ be the **marginal** densities. then

$$f(x,y) = f_X(x)f_Y(y|x) = f_Y(y)f_X(x|y), \quad x, y \in \mathbb{R}$$

$f_Y(\cdot|x)$ is the **conditional** density of Y given $X = x$
$f_X(\cdot|y)$ is the **conditional** density of X given $Y = y$

independence

X, Y are independent $\iff \forall x, y \in \mathbb{R}$,
1. $f(x,y) = f_X(x)f_Y(y)$
2. $f_Y(y|x) = f_Y(y)$
3. $f_X(x|y) = f_Y(x)$

X, Y are independent $\Rightarrow$
- $E(XY) = E(X)E(Y)$
- $\mathrm{cov}(X,Y) = 0$

(the converse does not hold)

Distributions

if X is iid, then $\mathrm{var}(\sum_{i=-1}^{n} x_i) = \sum_{i=1}^{n} \mathrm{var}(x_i)$

bernoulli

- $X \sim Bernoulli(p) \quad \Rightarrow$ coin flip with probability p

binomial

- $X \sim Bin(n,p) \quad \Rightarrow n$ coin flips with probability p
- $X_i \overset{i.i.d.}{\sim} Bernoulli(p)$
- $E(X) = \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$
  $E(X) = np, \quad \mathrm{var}(X) = np(1-p)$

multinomial

- $X \sim Multinomial(n, \mathbf{p}) \quad \Rightarrow n$ runs of an experiment with k outcomes with probability vector $\mathbf{p}$
  - An experiment with k outcomes $E_1, \dots, E_k$, $Pr(E_i) = p_i$. For some $1 \le i \le k$, let X_i be the number of times E_i occurs in n runs.

$(X_1, \dots, X_k)$ has the multinomial distribution:
$$Pr(X_1 = x_1, \dots, X_k = x_k) = \binom{n}{x_1 \dots x_k}\Pi_{i=1}^{k} p_i^{x_i}$$

- combinatorially, $\binom{n}{x_1 \dots x_k} = \frac{n!}{x_1! x_2! \dots x_k!}$

$$E(X) = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix}, \quad \mathrm{var}(X_i) = np_i(1-p_i)$$

$\mathrm{var}(X) =$ *covariance matrix* M with

$$m_{ij} = \begin{cases} \mathrm{var}(X_i) & \text{if } i = j \\ \mathrm{cov}(X_i, X_j) & \text{if } i \neq j \end{cases}$$

- $\mathrm{cov}(X_i, X_j) < 0$
- $X_i \sim Bin(n, p_i)$
  - $E(X_i) = np_i, \quad \mathrm{var}(X_i) = np_i(1-p_i)$
- $X_i + X_j \sim Bin(n, p_i + p_j)$
  - $\mathrm{var}(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j)$

Conditional expectation

discrete case

for r.v.s (X, Y), let $f_Y(\cdot|x_i)$ be the conditional mass function of Y given $X = x_i$.

$$E[Y|x_i] := \sum_{j=1}^{J} y_j f_Y(y_j|x_i)$$

$$\mathrm{var}[Y|x_i] := \sum_{j=1}^{J}(y_j - E[Y|x_i])^2 f_Y(y_j|x_i)$$

$E[Y|x_i]$ is like $E(Y)$, with conditional distribution replacing marginal distribution $f_Y(\cdot)$. likewise $\mathrm{var}[Y|x_i]$ is like $\mathrm{var}(Y)$

continuous case

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

$$\mathrm{var}[Y|x] := \int_{-\infty}^{\infty}(y - E[Y|x])^2 f_Y(y|x)\,dy$$

02. PROBABILITY (2)

mean square error (MSE)

mean square error, $MSE = E\{(Y-c)^2\}$

- $MSE = \mathrm{var}(Y) + \{E(Y) - c\}^2$
- Y and X are correlated:
  $MSE = \mathrm{var}[Y|x] + \{E[Y|x] - c\}^2$
  $MSE = E[(Y-c)^2|x] = E[\{Y - E(Y)\}^2|x]$
  - to predict Y, choose c that depends on x

random conditional expectations

let X, Y be r.v.s.
- $E[Y|X]$ is a r.v. which takes value $E[Y|x]$ with probability/density $f_X(x)$
- $\mathrm{var}[Y|X]$ is a r.v. which takes value $\mathrm{var}[Y|x]$ with probability/density $f_X(x)$
- $E(E[X_2|X_1]) = E(X_2)$
- $\mathrm{var}(E[X_2|X_1]) + E(\mathrm{var}[X_2|X_1]) = \mathrm{var}(X_2)$

mean MSE

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{var}[Y|x_i] \approx TODO$$

cumulative distribution function (cdf)

for r.v. X, let $F(x) = P(X \le x)$
- domain: $\mathbb{R}$; codomain: $[0,1]$

$$F(x) = \int_{-\infty}^{\infty} f(x)\,dx$$

standard normal distribution

$Z \sim N(0,1)$ has density function
$$\phi(z) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{z^2}{2}\}, \quad -\infty < z < \infty$$

- $E(Z) = 0, \quad \mathrm{var}(Z) = 1$
  - $E(Z) = \int_{-\infty}^{\infty} z\phi(z)\,dz$
  - $E(Z^2) = \int_{-\infty}^{\infty} z^2\phi(z)\,dz$
  - $E(Z^k) = \begin{cases} 0 & \text{if } k \text{ is odd} \\ 1 & \text{if } k \text{ is even} \end{cases}$
- CDF, $\Phi(x) = P(Z \le x), \ x \in \mathbb{R}$
  - $\Phi(x) = \int_{-\infty}^{x} \phi(z)\,dz$

general normal distribution

let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$

standardisation: $\frac{X-\mu}{\sigma} \sim N(0,1)$

- summations:
  - for constants $a, b \neq 0$,
    $a + bX \sim N(a + b\mu, \ b^2\sigma^2)$
  - $X + Y \sim N(\mu + \nu, \ \sigma^2 + \tau^2 + 2\,\mathrm{cov}(X,Y))$
    - $\mathrm{cov}(X,Y) = 0, \quad \Rightarrow \quad X \perp Y$
    - $X \perp Y \quad \Rightarrow \quad X + Y \sim N(\mu + \nu, \ \sigma^2 + \tau^2)$
- for $W = a + bX$,
  - density $f_W(w) = \frac{d}{dw}F_W(w)$
  - cdf $F_W(w) = P(X \le \frac{w-a}{b}) = \Phi(\frac{w-a}{b})$

Central limit theorem

let $X_1, \dots, X_n$ be iid rv's with expectation μ and SD σ, with $S_n \sum_{i=1}^{n} X_i$

CLT

as $n \to \infty$, the distribution of the standardised $S_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ converges to $N(0,1)$

- $E(S_n) = n\mu, \ \mathrm{var}(S_n) = n\sigma^2$
- for large n, approximately $S_n \sim N(n\mu, n\sigma^2)$

Bernoulli

let $X_i \sim Bernoulli(p)$. then
- $S_n \sim Binom(n,p)$
  - for large n, $S_n = N(np, np(1-p))$
- CLT: standardised $\frac{S_n - np}{\sqrt{n}\sqrt{p(1-p)}} \to N(0,1)$ as $n \to \infty$

χ^2 RVs

let $Z \sim N(0,1)$.

$Z^2 \sim \chi_1^2$
Z^2 has χ^2 distribution with 1 degree of freedom
$E(Z^2) = 1$
$\mathrm{var}(Z^2) = E(Z^4) - \{E(Z^2)\}^2 = 2$

let $V_1, \dots, V_n$ be iid χ_1^2 RVs. then
- $V = \sum_{i=1}^{n} V_i$ has a χ_n^2 distribution: $V \sim \chi_n^2$
- $E(V) = n \quad \mathrm{var}(V) = 2n$

Gamma distribution

let $\alpha, \lambda > 0$. The $Gamma(\alpha, \lambda)$ density is
$$\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, \quad x > 0$$
where $\Gamma(\alpha)$ is a number that makes density integrate to 1

- density of χ_1^2 RV $= \frac{1}{\sqrt{2\pi}}v^{-1/2}e^{-v/2}, \quad v > 0$
  $= Gamma(\frac{1}{2}, \frac{1}{2})$

- χ^2_n RV $\sim Gamma(\frac{n}{2}, \frac{1}{2})$
  - χ^2_n is a special case of Gamma!
- if $X_1 \sim Gamma(\alpha_1, \lambda)$ and $X_2 \sim Gamma(\alpha_2, \lambda)$ are independent, then $X_1 + X_2 \sim Gamma(\alpha_1 + \alpha_2, \lambda)$