ST2132 Probability Review I

Semester 1 2022/2023

If printing, do DOUBLE-SIDED, each side TWO slides.

Outline

- Basic concepts: random experiment, random variable, probability, realisation, distribution, expectation
- Additional concepts: standard deviation, variance, covariance, random vector, multinomial distribution, conditioning, Law of Large Numbers, Monte Carlo approximation

What is probability?

A **random experiment** can be repeated independently under identical conditions.

Numerical outcomes of a random experiment are represented by a **random variable** (RV).

The **probability** of an event is the limiting relative frequency of its occurrence as the experiment is repeated many times.

Mathematicians prefer to define probability and random variables using a sophisticated subject called “real analysis”. Our definition of probability is then a “frequency interpretation” of the mathematical definition.

Examples and non-examples

- Coin tossing: X is 0 (tail) or 1 (head). How do we know if the coin is fair:
$$\Pr(X = 1) = 0.5?$$
- Roulette: Y can be $0, 1, \ldots, 36$. How do we know if
$$\Pr(Y = i) = \frac{1}{37}?$$
- If no random experiment, hard to talk about probability. For instance, whether student e1234567 will have CAP above 4.5, or whether an accused is guilty as charged.
- Probability may apply for a large group of items. An epidemiologist may estimate the probability that a male senior citizen will be infected by COVID-19 in September 2022.

Realisations of a random variable

- Run experiment once to get x, a **realisation** of X.
- Run experiment n times to get realisations $x_1, \ldots, x_n$ of X.
- Unlike a random variable, realisations are fixed numbers. We try to distinguish with capital and small letters.
- What can you say about the mean of a large number of realisations of X?

Expectation of X

RV X has a **distribution**, described as follows.

- Discrete case: $\Pr(X = x_i) = p_i$, $i = 1, 2 \ldots, n$.

$$\mathrm{E}(X) = \sum_{i=1}^{n} x_i p_i$$

  n can be infinity, and yet $\mathrm{E}(X)$ can be finite.

- Continuous case: density function is $f(x), x \in \mathbb{R}$.

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$$

  In practice, we only need the interval where $f > 0$.

Other expectations

Let X be a continuous RV with density f.

Choose all that are equal to $\mathrm{E}(X^2)$.

(a) $\int_{-\infty}^{\infty} x^2 f(x^2)\,\mathrm{d}(x^2)$ (b) $\int_{-\infty}^{\infty} x^2 f(x^2)\,\mathrm{d}x$

(c) $\int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$ (d) $\int_{0}^{\infty} x^2 f(x^2)\,\mathrm{d}(x^2)$

(e) $\int_{0}^{\infty} x^2 f(x^2)\,\mathrm{d}x$ (f) $\int_{0}^{\infty} x^2 f(x)\,\mathrm{d}x$

For a function $h : \mathbb{R} \to \mathbb{R}$,

$$\mathrm{E}\{h(X)\} = \underline{\hspace{6cm}}$$

Manipulation of expectation

For RV X, let $\mu = \mathrm{E}(X)$. Show that:

(a) $\mathrm{E}(X - \mu) = 0$

(b) $\mathrm{E}\{X(X - \mu)\} = \mathrm{E}(X^2) - \mu^2$

What can you say about the spread of a large number of realisations of X?

Standard deviation and variance

The standard deviation of X is $\mathrm{SD}(X) = \sqrt{\mathrm{var}(X)}$, where the variance is defined by

$$\mathrm{var}(X) = \mathrm{E}\{(X - \mu)^2\}$$

Show that

$$\mathrm{var}(X) = \mathrm{E}(X^2) - \mu^2 \qquad (*)$$

Let c be a constant. How is $\mathrm{var}(X - c)$ related to $\mathrm{var}(X)$? Justify your answer intuitively, then check using (*) or the definition.

Summarising an RV

An RV X is completely described by its distribution. It is useful to summarise it with $\mu = \mathrm{E}(X)$ and $\sigma = \mathrm{SD}(X)$.

- We say

  "X is around μ, give or take σ or so."

  meaning: a realisation x is typically about σ away from μ.

- Apply to z, a realisation of $Z \sim$ Normal(0,1). What more can you say about many realisations $z_1, z_2, \ldots, z_n$?

Approximating expectations

Let $x_1, \ldots, x_n$ be realisations of X.

- Law of Large Numbers: As $n \to \infty$, it becomes more certain that the sample mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

is close to $\mathrm{E}(X)$.

- More generally, let h be a function. As $n \to \infty$, it becomes more certain that

$$\frac{h(x_1) + h(x_2) + \cdots + h(x_n)}{n}$$

is close to $\mathrm{E}\{h(X)\}$.

Computation

Two workhorses in R, Python, etc.:

- Simulate: generate lots of (pseudo)-realisations of a random variable.
- Loop: do the same calculation many times. (In R, some for-loops can be "vectorised".)

Monte Carlo approximation:

$$\mathrm{E}\{h(X)\} \approx \frac{h(x_1) + \cdots + h(x_n)}{n}$$

better as n gets larger.

If there is a formula for $\mathrm{E}\{h(X)\}$, use it to build your coding skill. When there is no formula, you will Monte Carlo to approximate it.

Let $X \sim \text{Binomial}(100, 0.8)$ produce realisations $x_1, \ldots, x_n$, where $n = 1000$. What is the value of the following, roughly?

$$\frac{x_1 + \cdots + x_n}{n}$$

$$\frac{x_1^2 + \cdots + x_n^2}{n}$$

$$\frac{\sqrt{x_1} + \cdots + \sqrt{x_n}}{n}$$

$$\frac{h(x_1) + \cdots + h(x_n)}{n}, \qquad h(x) = \begin{cases} 1, & x < 80 \\ 0, & x \geq 80 \end{cases}$$

Why not just simulate?

With suitable choices of h, all kinds of probabilities and expectations can be approximated by Monte Carlo.

Why bother with formulae?

- Faster and easier to compute.
- No issue with random error.

Still good to know probability very well.

To solve real problems, statisticians need to choose sensible random variables before computing. The art of choosing depends on both logic and scientific knowledge.

Independent and identically distributed (IID) RVs

We often denote n repeated trials of a random experiment by RVs $X_1, X_2, \ldots, X_n$. They are

- Independent. For instance,

$$\Pr(X_1 = a, X_2 = b) = \Pr(X_1 = a)\Pr(X_2 = b)$$

$$E(X_3 X_4 X_5) = E(X_3)E(X_4)E(X_5)$$

- Identically distributed: For any a, $\Pr(X_i = a)$ is the same across i.

IID RVs are fundamental to statistical practice.

Realisations of IID RVs

Let $x_1, \ldots, x_n$ be realisations of X. x_1 is around μ, give or take σ

or so; same for any other realisation.

- Is $x_1 + x_2$ around 2μ, give or take 2σ?
- Let $X_1, \ldots, X_n$ be IID RVs, where $X_i \stackrel{d}{=} X$ (equal in distribution).

  We can view x_i as a realisation of X_i, $1 \leq i \leq n$.

  $x_1 + x_2$ around ____________, give or take ____________.
- Let $\bar{x}$ be the mean of the x's. It is a realisation of

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

  $\bar{X}$ is around ____________, give or take ____________.

Fair coin toss

Toss a fair coin 100 times, independently and under the same conditions. Let y be the total number of heads. It is a realisation of a random variable Y.

- Write Y as a sum of IID RVs.
- Calculate $\mathrm{E}(Y)$ and $\mathrm{SD}(Y)$.
- y is about how far away from 50?
- $y/100$ is about how far away from 0.5?

Tabulating realisations

Arrange $10,000$ realisations from X in a 500×20 matrix.

- Each row is a realisation of which random vector?
- Each column is a realisation of which random vector?
- Each row sum is a realisation of which random variable?
- Each column sum is a realisation of which random variable?
- Does it matter how the arrangement is done?

Algebra of RV

Let X, Y be RVs, and a, b, c be constants.

- $Z = aX + bY + c$ is also an RV.
  If x and y are realisations of X and Y, then $ax + by + c$ is a realisation of Z. We may write $z = ax + by + c$.
- Can the density of Z be derived from f_X and f_Y, the respective densities of X and Y?
- The expectation is linear: Always

$$\mathrm{E}(Z) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y) + c$$

- Practically any function of RVs is an RV, subject to the usual constraint. For example, XY, e^{Y-X}. $\sqrt{X}$ and $\log X$ also, provided X is nonnegative.

Is a constant an RV?

Intuitively, the answer should be no.

- Which point in the previous slide implies that the answer is technically yes?
- Let H and T be the number of heads and tails respectively in 10 tosses of a coin. What is the distribution of their sum?
- Every theorem about an RV is true about a constant.

(Is a square a rectangle? Is $\emptyset$ a set?)

Two RVs

Let RVs (X, Y) have **joint** density $f(x, y)$.
For practically all cases,

$$f(x, y) = f_X(x) f_Y(y|x)$$

where
$f_X(x)$ is the **marginal** density of X,
$f_Y(y|x)$ is the **conditional** density of Y given $X = x$.
They are ordinary densities, the words in bold just referring to relationship with joint density.

Generally,

$$\text{joint} = \text{marginal} \times \text{conditional}$$

What happens if X and Y are independent?

Expectations

$\mathrm{E}(X) = \_\_\_\_\_$

(a) $\int_{-\infty}^{\infty} x\, f(x,y)\, \mathrm{d}x\mathrm{d}y$ (b) $\int_{-\infty}^{\infty} x\, f(x,y)\, \mathrm{d}x$

(c) $\int_{-\infty}^{\infty} x\, f_X(x)\, \mathrm{d}x\mathrm{d}y$ (d) $\int_{-\infty}^{\infty} x\, f_X(x)\, \mathrm{d}x$

$\mathrm{E}(XY) = \_\_\_\_\_$

(a) $\int_{-\infty}^{\infty} xy\, f(x,y)\, \mathrm{d}x\mathrm{d}y$ (b) $\int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y|x)\, \mathrm{d}x\mathrm{d}y$

(c) $\int_{-\infty}^{\infty} x\, f_X(x) (\int_{-\infty}^{\infty} y\, f_Y(y|x)\, \mathrm{d}y) \mathrm{d}x$

For a function $h : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathrm{E}\{h(X,Y)\} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

Covariance

Let $\mu_X = \mathrm{E}(X)$, $\mu_Y = \mathrm{E}(Y)$. The covariance between X and Y is

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

The covariance has no direct interpretation, but is a stepping stone to the correlation.

Show that $\mathrm{cov}(X, Y) = \mathrm{E}(XY) - \mu_X \mu_Y$. What is $\mathrm{cov}(X, X)$?

Always,

$$\mathrm{var}(aX + bY + c) = a^2\,\mathrm{var}(X) + b^2\,\mathrm{var}(Y) + 2ab\,\mathrm{cov}(X, Y)$$

- This game uses a fair die to simulate the annual weather, for 10 years.
- The outcomes are drought (1 spot), normal (2 to 5 spots), and flood (6 spots).
- Let D, N, F count the number of occurrences of the outcomes. What is the distribution of the random vector

$$\mathbf{X} = \begin{bmatrix} D \\ N \\ F \end{bmatrix}$$

Multinomial distribution

An experiment has r outcomes $E_1, \ldots, E_r$ with probabilities $p_1, \ldots, p_r$. For $1 \le i \le r$, let X_i be the number of times E_i occurs in n independent trials. $(X_1, \ldots, X_r)$ has the multinomial distribution:

$$\Pr(X_1 = x_1, \ldots, X_r = x_r) = \binom{n}{x_1 \cdots x_r} \prod_{i=1}^{r} p_i^{x_i}$$

where $x_1, \ldots, x_r$ are non-negative integers summing to n.

From a population with r kinds of individuals, make n random draws with replacement.

(a) What variables have a multinomial distribution?

(b) How can the probabilities be interpreted in the population?

Multinomial (2)

Let $r = 3$, so that $X = (X_1, X_2, X_3)$ has a trinomial distribution.

- Without calculation, decide whether X_1 and X_2 are positively correlated, negatively correlated, or uncorrelated.
- What is the distribution of X_1?
- What is the distribution of $X_1 + X_2$?
- State the expectation and variance of $\mathbf{X}$.

Conditional expectation

What can we say about the mean of a large number of realisations from the conditional distribution of Y given $X = x$?

Which is $\mathrm{E}[Y|x]$?

(a) $\int_{-\infty}^{\infty}(y|x)f_Y(y|x)\,\mathrm{d}(y|x)$ (b) $\int_{-\infty}^{\infty} y\, f_Y(y|x)\,\mathrm{d}(y|x)$

(c) $\int_{-\infty}^{\infty}(y|x)f_Y(y|x)\,\mathrm{d}y$ (d) $\int_{-\infty}^{\infty} y\, f_Y(y|x)\,\mathrm{d}y$

Officially, the conditional expectation is a new definition, but it should also be seen as a special case of an ordinary expectation.

$\mathrm{E}[Y|x]$ is like $\mathrm{E}(Y)$, but with a different distribution.

Conditional variance

What can we say about the spread of a large number of realisations from the conditional distribution of Y given $X = x$?

Express $\text{var}[Y|x]$ in two forms:

- As the expectation of something.
- As the sum of two terms.

Draws without replacement

A box has N identical tickets: K marked '1' and $N-K$ mark '0', where $0 < K < N$.

X_1, X_2: results of two random draws without replacement.

- What is the distribution of X_1?
- What is the conditional distribution of X_2 given $X_1 = 0$, and given $X_1 = 1$?
- Write down $\mathrm{E}[X_2|x_1]$, $\mathrm{var}[X_2|x_1]$, where $x_1 = 0, 1$.
- Rework the above, exchanging X_1 and X_2.

True or false?

$$\mathrm{E}[Y|X] = \int_{-\infty}^{\infty} (y|x) f(x,y)\, \mathrm{d}(y|x)$$

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

To understand these objects, you must become famliar with $\mathrm{E}[Y|x]$ and $\mathrm{var}[Y|x]$.

Conclusions

Revised key concepts and actions and how they fit together.

- Fundamental concepts: random experiment, probability, random variable, realisation, distribution, expectation
- Other concepts: variance, SD, covariance, random vector, multinomial distribution, conditioning, Law of Large Numbers
- Actions: simulation, Monte carlo approximation