Statistical Methods

Lecture 5 – Discrete Random Variables

Luke Dickens

Autumn 2020

What is a Random Variable?

Short Answer: A number determined by an experiment.
Motivation: Often with an experiment, we aren't interested in the precise outcome but in some numerical quantity determined by the outcome. Examples include:

- The sum on the faces of two rolled dice
  E.g. We don't care which colour dice shows a higher number.
- The money won from a specific bet on a horse race
  E.g. We care only whether our horse came first.
- The stock price of a given stock at a given time
  E.g. Not how much it fluctuates an hour before.
- The length of time it takes athlete X to run 100m on day Y
  E.g. Not whether she crosses the line left foot first.

But what is a Random Variable really?

Without going too deeply into the theory (measure theory). . .

Definition: For an experiment with sample space S, and a (well-behaved) probability distribution over S, a random variable (RV) is a (well-behaved) function from outcomes to real numbers, e.g. $X : S \to \mathbb{R}$.

Does a RV have to take real values?

Short Answer: *Yes. More generally, a function from outcomes to something else might be called a random quantity.*

Definition: A RV is said to be discrete if its possible values are separated points on the number line.

For instance:

- an RV that can take one of a finite number of different values
- an RV that can only take integer values

Examples:

- sum of shown faces of two rolled dice
- money won from a bet on a horse race
- number of defective items from a production line in an hour
- number of telephone calls received by a call centre on a day
- number of times you toss a tail before tossing a head

The length of time to run 100m is not a discrete RV.

Discrete RV: Notation

Let X be a discrete RV with n possible values, $x_1, x_2, \ldots, x_n$.
$P(X = x_i) \in [0, 1]$ represents the probability that X is equal to x_i.

The collection of probabilities, $P(X = x_i)$ for all i, is called the **probability distribution** of X (or probability mass function).

X must take one (and exactly one) of the n values x_i and so:

$$\sum_{i=1}^{n} P(X = x_i) = 1$$

We typically write $p(x_i) = P(X = x_i)$:

- $p : \{x_i\}_{i=1}^{k} \to [0, 1]$ is now a function
- argument indicates which RV, e.g. $p(x_i)$ for RV X and $p(y_j)$ for RV Y
- Safer to use $p_X(x_i)$ for X, and $p_Y(y_j)$ for Y

Cumulative Distribution Function

A related concept to the probability mass function is the cumulative distribution function (not discussed in [Ros17]).

Definition: For discrete RV X taking values $x_1, \ldots, x_k$ cumulative distribution function is the function F where:

$$F(x_i) = P(X \leq x_i) \quad \text{for } i = 1, \ldots, k$$

$F(x_i)$ is probability that X will take value less than or equal to x_i

- $F : \{x_i\}_{i=1}^k \to [0, 1]$ again a function
- Argument indicates RV (use F_X for safer notation)
- If $x_1, \ldots, x_k$ are ordered ($x_{i-1} \leq x_i$), then

$$F(x_{i-1}) \leq F(x_i) \qquad \text{and} \qquad F(x_k) = 1$$

Expected Value

If X is a discrete RV that takes on one of the possible values $x_1, x_2, \ldots, x_n$ then the expected value, expectation or mean of X:

$$E[X] = \sum_i x_i P(X = x_i)$$

Interpretation: Assume a very large number, N, of independent trials of an experiment are performed. Any event A with probability $P(A)$ will occur $P(A)$ of the time.

RV X taking values $x_1, x_2, \ldots, x_n$, with respective probabilities $p(x_1), p(x_2), \ldots, p(x_n)$; let's say X are winnings in single game of chance. Average winning per game will be

$$\frac{\sum_{i=1}^{n} x_i N p(x_i)}{N} = \sum_{i=1}^{n} x_i p(x_i) = E[X]$$

$p(x_i)$ is shorthand for $P(X = x_i)$

Example: Bernoulli Random Variable

Bernoulli random variable X takes values 1 or 0 with probabilities $P(X = 1) = p$ and $P(X = 0) = (1 - p)$. Expected value:

$$E[X] = p \cdot 1 + (1 - p) \cdot 0 = p$$

Interpretation: We can think of X as a numerical interpretation of the experiment of tossings a biased coin with:

- $S = \{\text{heads}, \text{tails}\}$
- $P(\{\text{heads}\}) = p$
- $P(\{\text{heads}\}^c) = P(\{\text{tails}\}) = (1 - p)$

Where the RV X maps outcome heads to 1 and tails to 0.

Single Die Example

RV X is value shown on single die, with probability distribution:

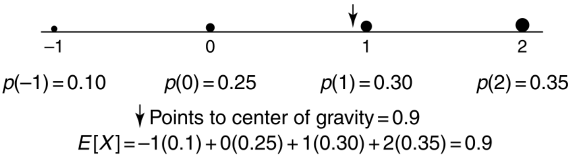
$$P(X = i) = \frac{1}{6} \quad \text{for } i = 1, \ldots, 6$$

Expected value:

$$E[X] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Note: $E[X]$ is not one value that we expect X to have, but rather the average value of X over a large number of repetitions.

Properties of expected value



From [Ros17], we have:

- expected value is analogous to centre of gravity of masses along a rod, where $P(X = x_i)$ is the *mass* of each point
- For constant c, $E[X + c] = E[X] + c$ and $E[cX] = cE[X]$
- For two RVs X and Y: $E[X + Y] = E[X] + E[Y]$
- More generally for k RVs $X_1, \ldots X_k$:

$$E[\sum_{i=1}^{k} X_i] = \sum_{i=1}^{k} E[X_i]$$

Example [Ros17, Ex. 5.11]

A contractor makes bids for 3 jobs. For each he wins he makes respective profits of \$20000, \$25000 and \$40000 with respective probabilities 0.3, 0.6 and 0.2. Any job he loses incurs a negative profit $-\$2000$. What is the expected profit.

X_i is the profit from the ith job and so:

- $E[X_1] = 0.3 \cdot \$20000 - 0.7 \cdot \$2000 = \$4600$
- $E[X_2] = 0.6 \cdot \$25000 - 0.4 \cdot \$2000 = \$14200$
- $E[X_3] = 0.2 \cdot \$40000 - 0.8 \cdot \$2000 = \$6400$
- Expected total profit is

$$E[X_1 + X_2 + X_3] = E[X_1] + E[X_2] + E[X_3] = \$25200$$

Note: Expected value has same unit as variable values

Variance of RVs

The expected value gives the long-term average of a RVs value, but tells us nothing about the variation, or spread of these values.

Definition: If X is a random variable with expected value $E[X] = \mu$, then the variance of X is

$$\text{var}(X) = E[(X - \mu)^2]$$

An alternative form is: $\text{var}(X) = E[X^2] - \mu^2$

Variance tells us about the variation/spread of our RV.

Properties of Variances

For constant c:

- $\text{var}(cX) = c^2\,\text{var}(X)$
- $\text{var}(X + c) = \text{var}(X)$

Definition: RVs X and Y are **independent** if knowing the value of one of them does not change the distribution of the other.

- For **two independent RVs**, X and Y,

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

- For k **independent RVs**, $X_1, \ldots, X_k$,

$$\text{var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} \text{var}(X_i)$$

Standard Deviation

Definition: The **standard deviation (SD)** is:

$$\mathrm{SD}(X) = \sqrt{\mathrm{var}(X)}$$

Like expected value, SD is measured in same units as the RV.

SD tells us about the variation/spread of our RV

For constant c:

- $\mathrm{SD}(cX) = |c|\,\mathrm{SD}(X)$
- $\mathrm{SD}(X + c) = \mathrm{SD}(X)$

For **two independent RVs**, X and Y,

$$\mathrm{SD}(X + Y) = \sqrt{\mathrm{var}(X + Y)} = \sqrt{\mathrm{var}(X) + \mathrm{var}(Y)}$$

Expectation of functions of RVs

Consider discrete RV X and function $g : \mathbb{R} \to \mathbb{R}$, then $g(X)$ is itself a random variable. (Remember any RV is just a function from outcomes to numbers.)

Expected value of $g(X)$ is:

$$E[g(X)] = \sum_{i=1}^{n} g(x_i) P(X = x_i)$$

This helps us to define $E[X^2]$ in:

$$\text{var}(X) = E[X^2] - E[X]^2$$

Examples in [Ros17, Exs. 5.16-5.18] demonstrate *the friendship paradox – on average your friends have more friends than you.*

Joint Distribution of RVs

Definition: Suppose X and Y are discrete random variables, with X taking values $x_1, \ldots, x_n$ and Y taking values $y_1, \ldots, y_m$. The probability that both $X = x_i$ and $Y = y_j$ is:

$$P(X = x_i, Y = y_j) \quad - \quad \text{shorthand} \quad p(x_i, y_j)$$

As outcomes, X and Y must take on exactly one of their respective values:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P(X = x_i, Y = y_j) = 1$$

As X can equal x_i in m mutually exclusive ways (one per value y_j, $j = 1, \ldots, m$), it follows that:

$$P(X = x_i) = \sum_{j=1}^{m} P(X = x_i, Y = y_j)$$

Definition: For two RVs X and Y their **covariance** is:

$$\mathrm{cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

where $\mu_x = E[X]$ and $\mu_y = E[Y]$

Equally: $\mathrm{cov}(X, Y) = E[XY] - \mu_x \mu_y$

Covariance measures degree to which the two RVs vary together:

- If $\mathrm{cov}(X, Y) > 0$, then large X $(> \mu_x)$ tend to be associated with large Y $(> \mu_y)$, and small X with small Y
- If $\mathrm{cov}(X, Y) < 0$, then large X tend to be associated with small Y $(< \mu_y)$, and small X with large Y
- captures direction and magnitude of how X and Y covary

Correlation

Definition: For two RVs X and Y their **correlation** is:

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} = \frac{\mathrm{cov}(X, Y)}{\mathrm{SD}(X)\,\mathrm{SD}(Y)}$$

Correlation measures degree to which the two RVs vary together:

- doesn't capture magnitude of variation
- **normalised** by SD of each variable
- lies between -1 and 1

Interpreted similarly to Pearson's r.

Related Concepts

Distributional measures relate to sample statistics:

Concept	Statistic	Measure
Centre	Sample Mean Sample Median	Expected Value Median*
Spread	Sample Variance Sample SD	Variance SD
Vary together	Sample Covariance* Pearson's r	Covariance Correlation

$*$ We did not introduce this concept

Binomial RVs

Definition [Ros17, Sec. 5.6]: n independent subexperiments are performed, each results in *success* with probability p or *failure* with probability $1 - p$. If X is the total number of successes, then X is a binomial RV with parameters n and p.

Written $X \sim \text{Binomial}(n, p)$, for $i = 0, \ldots, n$,

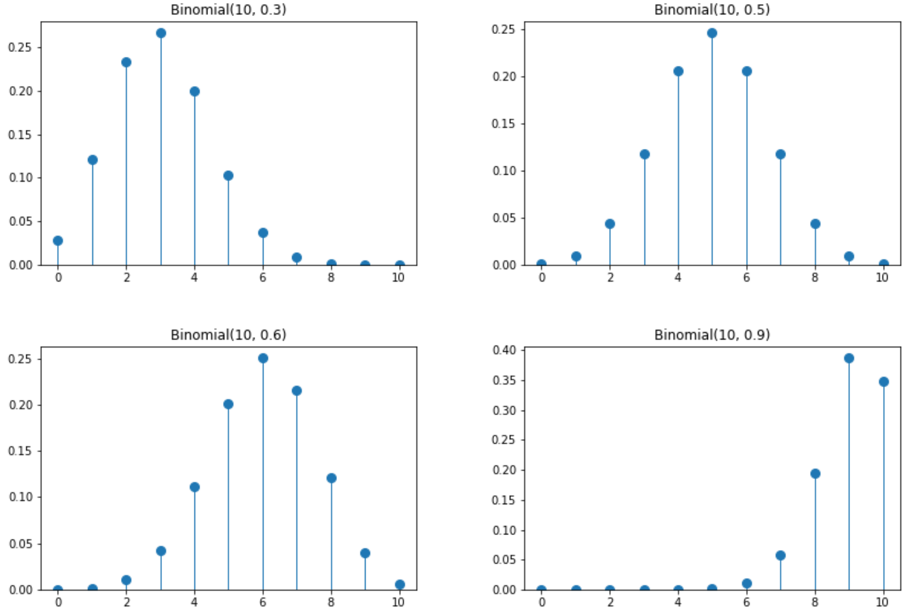
$$P(X = i) = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

Properties of Binomial RVs

For any RV $X \sim \text{Binomial}(n, p)$:

- $E[X] = np$
- $\text{var}(X) = np(1-p)$
- For $p = 0.5$, distribution is symmetric
- For large n and p not close to 0 or 1, X satisfies the empirical rule increasingly closely (as we will see).

Visualising the Binomial



Example [Ros17, Ex. 5.24]

When two individuals mate, the child gets one gene from each parent; this gene is equally likely to be either of the parent's two genes. Brown eye genes are dominant, blue eyes recessive: you need two blue eye genes to be blue eyed. Consider two hybrid parents, with 1 brown and 1 blue eye gene.

(a) With what probability will their child have blue eyes?

Both parents donate a blue-eye gene is event B then:

$$P(\mathit{blue\ eyes}) = P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

(b) Of 4 children, with what probability will one be blue-eyed?

Number of children with blue eyes is RV $X \sim \text{Binomial}(4, \frac{1}{4})$.

$$P(X = 1) = \binom{4}{1} P(B)^1 P(B^c)^3 = \frac{4!}{3!1!} \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^3 = \frac{27}{64}$$

Hypergeometric RVs

Definition [Ros17, Sec. 5.7]: n items are randomly selected from N possibilities without replacement, of which $K = Np$ represent *successes* and the other $N - K = N(1 - p)$ represent *failures*. If X is equal to the number of *successes* in the sample, then X is a hypergeometric RV with parameters n, N and K (or p).

Written $X \sim \text{Hypergeometric}(N, K, n)$, the probability of drawing exactly k successes is $P(X = k)$

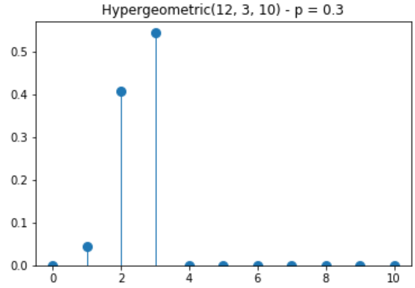
$$= \begin{cases} \dfrac{\dbinom{K}{k}\dbinom{N-K}{n-k}}{\dbinom{N}{n}} & \text{if } \max(0, n+K-N) \leq k \leq \min(K, n) \\ 0 & \text{otherwise} \end{cases}$$

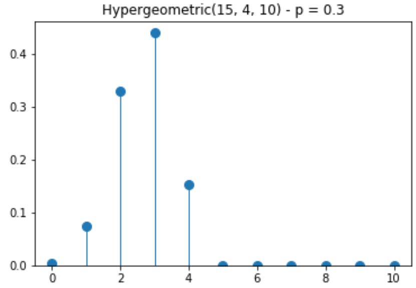
Properties of Hypergeometric RVs

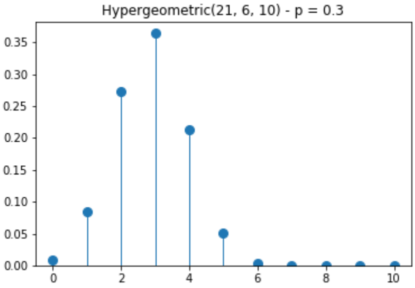
Consider an RV $X \sim \text{Hypergeometric}(N, K, n)$ with $p = \frac{K}{N}$. And for comparison, another RV $Y \sim \text{Binomial}(n, p)$:

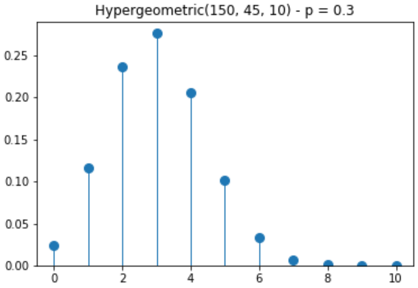
- $E[X] = n\frac{K}{N} = np = E[Y]$
- $\text{var}(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1} = np(1-p)\frac{N-n}{N-1}$
- $\text{var}(X) < np(1-p) = \text{var}(Y)$
- draws (subexperiments) for hypergeometric are not identically distributed
- For large N and much smaller n, X is distributed very similarly to RV Y

Visualising the Hypergeometric









Poisson RVs

Definition [Ros17, Sec. 5.8]: RV X is called a Poisson with parameter $\lambda > 0$, written $X \sim \text{Poisson}(\lambda)$ if for $i = 0, 1, \ldots$

$$P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}$$

With properties:

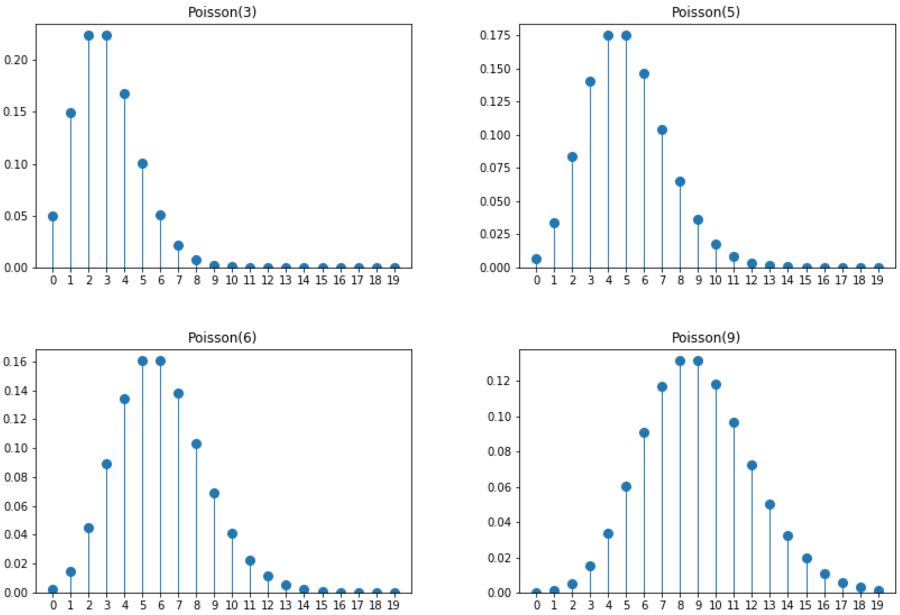
- $E[X] = \lambda$
- $\text{var}(X) = \lambda$

Poisson as an approximate binomial

Intuition [Ros17, Sec. 5.8]: Poisson RV can we seen as an approximation to binomial RVs. Consider n independent trials, each with probability of success p, If n is large and p then the total number of succceses will be approxiately poisson with $\lambda = np$.
Examples:

- Number of misprints on a page of a book
- Number of people in community who are over 100
- Number of people entering a post office on a given day

Visualising the Poisson



References I

[Ros17] Sheldon M. Ross, *Introductory Statistics*, 4 ed., Academic Press, 2017.