Goals of this section

Define:

- *random variables*.
- *discrete* random variables.
- *Bernoulli*, *Binomial*, and *Poisson* random variables.
- *Probability Mass Function* (pmf) or *probability distribution*.
- *Cumulative Distribution Function* (CDF).

- *Expected Value*.

- *Mean*, *Variance*, *Standard Deviation*.

- *Poisson processes*.

Random Variables

- A random variable is a numerical value determined by the outcome of a random experiment.

- An *rv* is a function whose domain is $\mathcal{S}$, the sample space, and whose range is a subset of the real numbers.

- Sometimes called a *real random variable* (to allow for things like random vectors, random complex numbers, random matrices, random functions and so on).

Example: Toss a pair of dice (one red, one green).

- Outcome is (r, g).
- Each of r and g is in $\{1, \ldots, 6\}$.
- X is total number of spots: $X(r, g) = r + g$.
- For three coin tosses. X is number of heads. Table of values

Outcome	HHH	HHT	HTH	HTT
X	3	2	2	1
	THH	THT	TTH	TTT
X				

Discrete Random Variables

- We describe some events by writing things like $X = 2$ or $X < 1$.

- Three coin tosses:

$$\begin{aligned}\{X = 2\} &= \{s \in \mathcal{S} : X(s) = 2\} \\ &= \{HHT, HTH, THH\}\end{aligned}$$

- Throw two dice:

$$\{X = 7\} = \\ \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

- A random variable is discrete if there is a set $\{x_1, x_2, \cdots\}$ such that

$$P(X \in \{x_1, x_2, \cdots\}) = 1.$$

- Distinguished from *continuous* rvs.

- Apply axioms of probability.

- First write

  $$\{s : X(s) \in \{x_1, x_2, \cdots\}\} = \cup_{j=1}^{\infty}\{s : X(s) = x_j\}$$

- These events are mutually exclusive.

Probability Mass Function

- So

$$P(X \in \{x_1, x_2, \cdots\}) = \sum_{j=1}^{\infty} P(X = x_j) = 1.$$

- Define the *probability mass function (pmf)* of X as:

$$p_X(x) = p(x) = P(X = x) = P(\{s : X(s) = x\}).$$

- Notice subscript X to say which rv *if needed*.

- Pay attention to two properties:
  - For any x we have $p_X(x) \geq 0$.
  - Law of total probability: $\sum_x p_X(x) = 1$.
- Usually $\{x_1, x_2, \cdots\}$ is a set of integers.
- It might be finite in spite of the $\cdots$.

Example: Bernoulli pmf

- Define *Bernoulli* rv: X has a Bernoulli(α) distribution if
  - Possible values of X are 0 and 1.
  - $P(X = 1) = \alpha$ and $P(X = 0) = 1 - \alpha$.
- The quantity α is a *parameter* of this distribution.
- **Example**: toss coin once. X is number of heads (in 1 toss!) so X is either 1 or 0.
- For fair coin $\alpha = 1/2$.
- Throw die. X is number of 6s. What is α?

Example: Another pmf

- Toss 2 dice, X is sum.

- Find pmf of X.

Example: A third pmf

- Toss 3 coins, X is number of Heads.

- Find pmf of X.

Example: A fourth pmf

- Throw pair of dice until you get a 4.

- Find pmf of X =number of throws till you get a 4.

Cumulative distribution functions

- For any rv X the cumulative distribution function (cdf) is

$$F_X(x) = P(X \leq x).$$

- Richard will graph cdf of number of 6s in 1 toss of fair die.

- Richard will graph cdf of # Heads in 3 tosses.

Properties of cdfs

- For $x < 0$ we have $F_X(x) = P(X \leq x) = 0$.

- In general for discrete X:

$$F(x) = \sum_{\{y: y \leq x\}} p(y)$$

- Basic features:

  - 0 at $-\infty$, goes to 1 at $+\infty$:

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to +\infty} F(x) = 1$$

  - F is non-decreasing.

  - F is right continuous and has left limits:

$$F(x) = \lim_{y \to x+} F(y) \text{ and } \lim_{y \to x-} F(y) \text{ exists.}$$

Example: Roll 8 sided die. Get number from 1 to 8. What is $F(6.5)$?

Expected Values

- For discrete rv X: *expected value* or *expectation* or *mean* of X is

$$\mathsf{E}(X) = \mu_X = \sum_x x \cdot p(x).$$

- Sum over x is over all possible values of X.

- For a Bernoulli(α) rv the expected value is $1\alpha + 0(1-\alpha) = \alpha$.

- Throw 2 dice. X is total showing.

$$\mathsf{E}(X) = 2\frac{1}{36} + 3\frac{2}{36} + \cdots + 7\frac{6}{36} + \cdots + 12\frac{1}{36} = 7.$$

The law of the unconscious statistician

- Expected value of $Y = (X - 7)^2$.
- Possible vals of Y: 0, 1, 4, 9, 16, 25.
- Probs are 6/36, 10/36, 8/36, 6/36,4/36, 2/36.
- So expected value is

$$\frac{10 + 4 \cdot 8 + 9 \cdot 6 + 16 \cdot 4 + 25 \cdot 2}{36} = \frac{210}{36}.$$

- Notice that, e.g.,

$$10/36 = P(Y = 1) = P(X = 8) + P(X = 6).$$

Simplified formulas

- That calculation of $\text{E}(Y)$ is
$$\text{E}(Y) = \sum_y y p(y)$$
but you could do a sum over x instead, as follows.

- Suppose $Y = h(X)$ for some function h. (In example $h(x) = (x - 7)^2$.)

- Then
$$\begin{aligned}\sum_y y P(Y = y) &= \sum_y y p_Y(y) \\ &= \sum_x h(x) P(X = x) \\ &= \sum_x h(x) p_X(x).\end{aligned}$$

Rules for expectation

- Expected value of a constant is that constant.

- That is
$$\mathrm{E}(a) = a$$
(For instance $\mathrm{E}(17) = 17$.)

- Expected values are *linear*.

- Meaning: if X and Y are rvs and a and b are constants then
$$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y).$$

Variances and Standard Deviations

- The variance of a random variable X is

$$\begin{aligned}\mathsf{E}\left((X-\mu_X)^2\right) &= \mathsf{E}\left(X^2 - 2\mu_X X + \mu_X^2\right)\\ &= \mathsf{E}(X^2) - 2\mu_X \mathsf{E}(X) + \mathsf{E}(\mu_X^2)\\ &= \mathsf{E}(X^2) - \mu_X^2\\ &= \mathsf{E}(X^2) - \mathsf{E}^2(X).\end{aligned}$$

- The units of the variance of X are the square of the units of X

- The Standard Deviation of X is

$$\sigma_X = \sqrt{\mathsf{Var}(X)}$$

- The SD has units the same as those of X.

- Statisticians do science a disservice when they forget the units.

- The standard deviation of adult male heights is 3 is not useful.

- Basic property:

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

  and the SD of $aX + b$ is $|a|$ times the SD of X.

The Binomial distribution

- Repeat some basic experiment a fixed number of times; call this number n.

- Each time we look to see whether or not some particular thing has happened.

- Label a trial a "Success" (S) if the "thing" happens.

- Otherwise a "Failure" (F).

- Let X be the number of successes.

- If we assume
  - The trials are independent.
  - The probability p of Success on a given trial is the same for every trial.

  then X has a Binomial(n, p) distribution.

The Binomial pmf

- Formula for the pmf:

$$
\begin{aligned}
p_X(k) &= P(X = k) \\
&= \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots, n
\end{aligned}
$$

- Probability of any particular sequence of k Ss and $n - k$ Fs is

$$p^k (1-p)^{n-k}$$

- The Binomial coefficient $\binom{n}{k}$ counts how many ways to arrange k Ss and $n - k$ Fs.

Mean, variance and SD of Binomial

- Suppose X is Binomial(n, p). Then

$$\begin{aligned} \mathsf{E}(X) &= np \\ \mathsf{Var}(X) &= np(1-p) \\ \sigma_X &= \sqrt{np(1-p)}. \end{aligned}$$

- Useful trick. Let X_i be 1 if trial i is a success and 0 if trial is a failure.

- Counting is adding:

- So

- Expected value of individual X_i:

- So

- Later use same trick on variance.

Poisson distribution

- Used as model for rare events.
- Number of atoms of Uranium in ore sample decaying in next second.
- Might seem better to use Binomial model with huge n and tiny p.
- But then must work with big powers of numbers a tiny bit less than 1.
- Actually take limit as $n \to \infty$ and $np \to \mu$.
- Get new distribution:

$$P(X = k) = \frac{\mu^k}{k!} e^{-\mu} \quad k = 0, 1, \ldots$$

- Mean and Variance of Binomial were

$$np \to \mu$$

and

$$np(1-p) \to \mu$$

so

$$\mathsf{E}(X) = \mathsf{Var}(X) = \mu.$$

- I am assigning you to read about the Poisson process.