Ex 3.14 (The gambler's ruin problem). Two gamblers, play the game of "heads or tails," in which each time a fair coin lands heads up player A wins \$1 from B, and each time it lands tails up, player B wins \$1 from A. Suppose that player A initially has a dollars and player B has b dollars. If they continue to play this game successively, what is the probability that (a) A will be ruined; (b) the game goes forever with nobody winning?

Sol: (The first step analysis). Let

$$\begin{aligned} E &= \{\text{ A (with i units to start) ruins}\} \\ H &= \{\text{ the first flip is head}\}. \end{aligned}$$

Then in the case $\mathbb{P}(H) = p$, $\mathbb{P}(H^c) = q$, $p + q = 1$,

$$\begin{aligned} p_i = \mathbb{P}(E) &= \mathbb{P}(E|H) \cdot \mathbb{P}(H) + \mathbb{P}(E|H^c) \cdot \mathbb{P}(H^c) \\ &= p_{i+1} \cdot p + p_{i-1} \cdot q \end{aligned}$$

with boundary values $p_0 = 1$ and $p_{a+b} = 0$. **Why?**

•A general principle to find the boundary conditions: Inside the recursive relation.

To solve the relation, we note

$$p(p_{i+1} - p_i) = q(p_i - p_{i-1})$$

and thus

$$p_{i+1} - p_i = (q/p)(p_i - p_{i-1}) = \cdots = (q/p)^i(p_1 - p_0)$$

Add the above for $i = 1, \cdots, j-1$, we have

$$p_j = \begin{cases} p_0 + (p_1 - p_0)(1 - (q/p)^j)p/(p-q) & \text{for} \quad p \neq q \\ ??? & \text{for} \quad p = q = 1/2. \end{cases}$$

The value of p_1 can be found by taking $j = a + b$. The final answer is ...

• The expected number of play is ab for $p = q = 1/2$ and can be easily found by using conditioning expectation.

Random Variables

Definition: A random variable (r.v.) is a real-valued function on the sample space S, i.e. $X : S \to \mathbb{R}$.

Ex: Toss a fair coin three times. Let X be the number of heads. Then possible values for X are $0, 1, 2, 3$, and

$$\begin{aligned}
\mathbb{P}(X = 0) &= 1/8 \\
\mathbb{P}(X = 1) &= 3/8 \\
\mathbb{P}(X = 2) &= 3/8 \\
\mathbb{P}(X = 3) &= 1/8
\end{aligned}$$

Thus $\mathbb{P}(\text{at least 2 heads}) = \mathbb{P}(X \geq 2) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = 1/2$. $\mathbb{P}(\text{at most 2 heads}) = \mathbb{P}(X \leq 2) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 7/8$

HW: S4.2: 1, 3, **4**, 5, **6**, 7; S4.3: 3, **4**, 5, **8**, 11: S4.4: 3, 5, **6**, 9, 11, **12**; S4.5: 3, **4**, 5, 7, **8**; S4.6: **2**; R4: 1, 5, **6**, 9, **10**.

Ex: Select at random 3 balls from an urn with 3 white, 3 red, 5 black. We win a dollar for each white ball selected and loss a dollar for each red ball selected. Let X be total winnings. Then X can only take values $0, \pm 1, \pm 2, \pm 3$, and

$$\begin{aligned}
\mathbb{P}(X=0) &= \frac{\binom{5}{3} + \binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}} \\
\mathbb{P}(X=1) = \mathbb{P}(X=-1) &= \frac{\binom{3}{1}\binom{5}{2} + \binom{3}{2}\binom{3}{1}}{\binom{11}{3}} \\
\mathbb{P}(X=2) = \mathbb{P}(X=-2) &= \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} \\
\mathbb{P}(X=3) = \mathbb{P}(X=-3) &= \frac{\binom{3}{3}}{\binom{11}{3}}
\end{aligned}$$

Note that $\mathbb{P}(\cup_{i=-3}^{3}\{X=i\}) = 1$ and

$$\mathbb{P}(\text{win money}) = \mathbb{P}(X>0) = \sum_{i=1}^{3} \mathbb{P}(X=i) = 1/3.$$

Distribution Functions

The (cumulative) distribution function $F(x)$ or $F_X(x)$ of the random variable X is

$$(cdf) \qquad F(b) = \mathbb{P}(X \le b) \qquad -\infty < b < \infty$$

Basic properties of c.d.f. $F(b)$:

- $F(x)$ is a nondecreasing function, i.e. if $a < b$, then $F(a) \le F(b)$.
- $\lim_{b\to\infty} F(b) = 1$ and $\lim_{b\to-\infty} F(b) = 0$.
- $F(x)$ is right continuous, i.e. $\lim_{b_n \to b^+} F(b_n) = F(b)$.

Note that $\mathbb{P}(X \ge b) = 1 - \mathbb{P}(X < b)$, $\mathbb{P}(a < X \le b) = F(b) - F(a)$, $\mathbb{P}(X > a) = 1 - F(a)$, but $\mathbb{P}(X < b) \ne F(b)$ in general.

Ex: Toss a fair coin three times. Let X be the number of heads.

$$F(x) = \mathbb{P}(X \le x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 1/2 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

Examples: 4.7, 4.8, 4.10.

Discrete Random Variable

A random variable that can take at most countable number of values is said to be a *discrete random variable*. For a discrete r.v. X, we define the probability mass function $p(a)$ or $p_X(a)$ of X by

$$p(a) = p_X(a) = \mathbb{P}(X = a).$$

If X can only take values $x_1 < x_2 < \cdots < x_n < \cdots$, then $p(x_i) \geq 0$, $i = 1, 2, \cdots$, $p(x) = 0$ for all other value of x, and $\sum_{i=1}^{\infty} p(x_i) = 1$. The c.d.f. of X is

$$F_X(a) = \mathbb{P}(X \leq a) = \sum_{\text{all} x \leq a} p(x).$$

Note that $F_X(x)$ is a step function with a step of size $p(x_i)$ at x_i, $i = 1, 2, \cdots$.

Previous Ex: $p(0) = 1/8$, $p(1) = 3/8$, $p(2) = 3/8$, $p(3) = 1/8$.

Expected value of discrete r.v. X

The *expectation* or *expected value* of a discrete r.v. X is

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x:p(x)>0} x \cdot p(x) \\ &= \text{Sums of value times probability}\end{aligned}$$

This is the weighted average of possible values of X, or center of gravity. The expectation of a r.v. is one of the most important concepts in probability.

Previous Ex: $\mathbb{E}\,X = \sum_{i=0}^{3} i \cdot p(i) = 0 \cdot 1/8 + 1 \cdot 3/8 + 2 \cdot 3/8 + 3 \cdot 1/8 = 3/2$.

Ex: The function I_A is an indicator variable for the event A if

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A^c \text{ occurs} \end{cases}$$

Then

$$\mathbb{E}\,I_A = 0 \cdot \mathbb{P}(I_A = 0) + 1 \cdot \mathbb{P}(I_A = 1) = \mathbb{P}(A).$$

Expectation of a function of a r.v.

Ex: Given the prob. mass function

$$\mathbb{P}(X=-1)=0.2, \mathbb{P}(X=0)=0.5, \mathbb{P}(X=1)=0.3.$$

Compute $\mathbb{E}\,X^2$.

Sol. 1: Let $Y=X^2$. Then the p.m.f. of Y is

$$\mathbb{P}(Y=0)=0.5, \quad \mathbb{P}(Y=1)=0.5.$$

Thus $\mathbb{E}\,Y=0\cdot 0.5+1\cdot 0.5=0.5$.

Sol. 2: The expectation is just sums of value times probability. So

$$\begin{aligned}\mathbb{E}\,X^2 &= (-1)^2\cdot\mathbb{P}(X=-1)+0^2\cdot\mathbb{P}(X=0)+1^2\cdot\mathbb{P}(X=1)\\ &= (-1)^2\cdot 0.2+0\cdot 0.5+1^2\cdot 0.3=0.5.\end{aligned}$$

Note that $\mathbb{E}\,X^2\neq(\mathbb{E}\,X)^2$.

Theorem 4.2: If X is a discrete r.v. that takes on one of the values x_i, $i\geq 1$, with respective probability $p(x_i)$, then for any function $g(x)$,

$$\begin{aligned}\mathbb{E}\,(g(X)) &= \sum_i g(x_i)p(x_i)\\ &= \text{sums of value times probability}\end{aligned}$$

Ex: The n-th moment of X is $\mathbb{E}\,X^n=\sum_i x_i^n p(x_i)$.

Ex 4.23: Assume $\mathbb{P}(X=i)=p(i)=i/15$, $i=1,2,3,4,5$. What is the expected value of $X(6-X)$?

Ex 4.23': What is the expected value of $|X-3|$?

Linearity of expectation: If a ad b are constants, then

$$\mathbb{E}\,(aX+b)=a\mathbb{E}\,X+b.$$

Proof: Assume $\mathbb{P}(X=i)=p(i)$, then

$$\begin{aligned}\mathbb{E}\,(aX+b) &= \sum_i (ai+b)p(i)\\ &= a\sum_i ip(i)+b\sum_i p(i)\\ &= a\mathbb{E}\,X+b.\end{aligned}$$

Ex 4.23": What is the expected value of $X^2-2|X|+1$?

Other examples: 4.15; 4.16; 4.17; 4.24.

Variance of R.V.

The variance of a r.v. X is defined by

$$\text{Var}(X) = \mathbb{E}\,(X - \mu)^2, \quad \text{where} \quad \mu = \mathbb{E}\,X.$$

It is a measurement of the spread of X around its mean $\mu = \mathbb{E}\,X$. The standard deviation is $\text{SD}(X) = \sqrt{\text{Var}(X)}$.

Basic properties of Var(X):

- $\text{Var}(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2$.
- For any $a, b \in \mathbb{R}$, $\text{Var}(aX + b) = a^2\text{Var}(X)$.
- For any $a \in \mathbb{R}$,

$$\mathbb{E}\,(X - a)^2 \geq \mathbb{E}\,(X - \mu)^2 = \text{Var}(X).$$

Ideas of Proofs: Definition and then simple algebra.

Ex 4.27: What is the variance of the random variable X, the outcome of rolling a fair die?

Ex 4.28: Suppose that, for a discrete random variable X, $\mathbb{E}\,X = 2$ and $\mathbb{E}\,(X(X-4)) = 5$. Find the variance and the standard deviation of $-4X + 12$.

Sure Betting on Different Beliefs

Alice believes that Republicans will win the election with probability 5/8. Bob believes that Democrats will win the election with probability 3/4. Assuming that Alice and Bob are both willing to accept any bet that gives them a positive expectation of winning, did you know that there's a way to place bets with both of them so that you can make money for certain?

Here's what you can do. Bet with Alice that you'll pay her 2 if Republicans wins and she'll pay you 3 otherwise. Alice agrees because her expectation is: $2(5/8) - 3(3/8) = 1/8$. Bet with Bob that you'll pay him 2 if Democrats wins, and he'll pay you 3 otherwise. Bob agrees because his expectation is $2(3/4) - 3(1/4) = 3/4$.

In general, as long as Alice and Bob have different beliefs about the probability of the outcomes of the election, you can design a bet that will give both of them positive expectation and you positive winnings!

Ref: K. Binmore, Fun and Games: a Text on Game Theory, 1992.

The Bernoulli and Binomial r.v's

The Bernoulli r.v. takes only two values 0 and 1 with

$$p(1) = \mathbb{P}(X = 1) = p, \quad p(0) = \mathbb{P}(X = 0) = 1 - p$$

where p is the probability of success. Note that $\mathbb{E}\, X = p$ and $\text{Var}(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = p(1 - p)$.

Suppose that n independent trials, each of which results in a success with probability p and in a failure with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a *binomial* random variable with parameters (n, p). we also write $X \sim \text{bi}(n, p)$ to represent binomial r.v. The probability mass function is given by

$$p(i) = \mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad 0 \le i \le n.$$

HW: S5.1: 1, 3, 9, 15, **20**, **26**; S5.2: 1, 3, 5, **6**, **8**; S5.3: 3, **4**, 7, 9, 15, **16**; R5: 1, 3, 5, **14**, **18**, **22**, 23.

• The binomial sum

$$\sum_{i=0}^{n} \mathbb{P}(X = i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = 1.$$

• The mean

$$\mathbb{E}\, X = \sum_{i=0}^{n} i \cdot \mathbb{P}(X = i) = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = np.$$

• The variance

$$\mathrm{Var}(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = np(1-p).$$

• The mass function $\mathbb{P}(X = k)$ is increasing for $0 \le k \le [(n+1)p]$, and is decreasing for $[(n+1)p] \le k \le n$. The proof is based on the fact that the ratio

$$\frac{P(X = k)}{\mathbb{P}(X = k-1)} = \frac{(n-k+1)p}{k(1-p)} \le 1 \quad \text{iff} \quad k \le (n+1)p.$$

EX: 5.1; 5.2; 5.3; 5.6; 5.7; 5.9.

The Poisson random variable

A r.v. X, taking on one of the values $0, 1, 2, \cdots$, is said to be a *Poisson* r.v. with parameter λ if for some $\lambda > 0$,

$$p(i) = \mathbb{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \cdots .$$

- The sum $\sum_{i=0}^{\infty} \mathbb{P}(X = i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$.
- The mean

$$\mathbb{E}\, X = \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i) = e^{-\lambda} \sum_{i=0}^{\infty} i \cdot \frac{\lambda^i}{i!} = \lambda.$$

- The variance

$$\mathrm{Var}(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

- **Poisson Approximation:** Let $X \sim \mathrm{bi}(n, p_n)$ and $\lambda = n \cdot p_n$ fixed. Then

$$\mathbb{P}(X = i) \approx e^{-\lambda}\frac{\lambda^i}{i!} \quad \text{for } n \text{ large.}$$

Note that p_n is small for n large.

EX: 5.10; 5.11; 5.12; 5.13.

The Geometric random variable

Suppose that independent trials each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs. Then the number of trials required, X, have the geometric distribution

$$p(n) = \mathbb{P}(X = n) = (1-p)^{n-1}p, \quad n = 1, 2, \cdots .$$

- The sum

$$\sum_{n=1}^{\infty} \mathbb{P}(X = n) = \sum_{n=1}^{\infty} (1-p)^{n-1}p = 1.$$

- The mean

$$\mathbb{E}\, X = \sum_{n=1}^{\infty} n \cdot \mathbb{P}(X = n) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = \frac{1}{p}.$$

- The variance

$$\mathbb{E}\, X = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = \frac{1-p}{p^2}.$$

- The memoryless property:

$$\mathbb{P}(X = m + n \,|\, X > m) = \mathbb{P}(X = n) \quad m, n \geq 1.$$

Ex: 5.18; 5.19.

The negative binomial r.v.

Suppose that independent trials each having a probability p, $0 < p < 1$, of being a success, are performed until a total of r successes is accumulated. Then the number of trials required, X, have the negative binomial distribution

$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \cdots .$$

- The sum

$$\sum_{n=1}^{\infty} \mathbb{P}(X = n) = \sum_{n=1}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = 1.$$

- The mean

$$\mathbb{E}\, X = \sum_{n=1}^{\infty} n \cdot \binom{n-1}{r-1} p^r (1-p)^{n-r} = \frac{r}{p}.$$

- The variance

$$\mathbb{E}\, X = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = \frac{r(1-p)}{p^2}.$$

Ex: 5.20.

• The negative binomial r.v. X can be represented as

$$X = Y_1 + Y_2 + \cdots + Y_r$$

where Y_1 equals the number of trials required for the first success, Y_2 the number of additional trials after the first success until the second success occurs, and Y_j the number of additional trials after the $(j-1)$th success until the jth success occurs, $1 \le j \le r$. Note that Y_j are ind. geometric r.v's. Thus

$$\mathbb{E}\,X = \mathbb{E}\,(Y_1 + \cdots + Y_r) = \mathbb{E}\,Y_1 + \cdots + \mathbb{E}\,Y_r = \frac{r}{p}$$

and

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(Y_1 + \cdots + Y_r) \\
&= \mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_r) \\
&= \frac{r(1-p)}{p^2}.
\end{aligned}$$

See Ex. 10.7 and 10.16 in Chapter 10 for details.

The Hypergeometric r.v.

A sample of n balls is chosen randomly (without replacement) from an urn containing N items, of which D are defective and $N-D$ are nondefctive. Let X denote the number of defective items selected. Then

$$\mathbb{P}(X=i)=\frac{\binom{D}{i}\binom{N-D}{n-i}}{\binom{N}{n}},\quad i=0,1,\cdots,n.$$

and X is called a *hypergeometric* r.v.

• The range of X is from $n-(N-D)$ to $\min(D,n)$ and the above mass function is always valid due to convention that $\binom{r}{k}=0$ when either $k<0$ or $r<k$.

• $\mathbb{E}\,X=nD/N$.

Ex: 5.23, 5.25.

Ex: Maximum Likelihood Method for parameter estimation. An unknown number, say N, of animals inhabit a certain region. To obtain some information about the population size, ecologies often perform the following experiment: They first catch a number, say m, of these animals, mark them in some manner, and release them. After allowing the marked animals time to disperse throughout the region, a new catch of size, say n, is made.

Let X denote the number of marked animals in this second capture. If we assume that the population of animals in the region remained fixed between the time of the two catches and that each time an animal was caught it was equally likely to be any of the remaining uncaught animals, it follows that X is a hypergeometric random variable such that

$$\mathbb{P}(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \equiv P_i(N)$$

Suppose not that X is observed to equal i. Then, as $P_i(N)$ represents the probability of the observed event when there are actually N animals in the region, it would appear that a reasonable estimate of N would be the value of N that maximizes $P_i(N)$. Such an estimate is called a *maximum likelihood* estimate (MLE). The maximization of $P_i(N)$ can be found by using

$$\frac{P_i(N)}{P_i(N_1)} = \frac{(N-m)(N-n)}{N(N-m-n+i)}.$$

The ratio is greater than 1 iff $N \leq mn/i$. So the MLE for N is $[mn/i]$, the largest integer not exceeding mn/i.