ST 501: Intro Probability & Distribution Theory

Discrete Random Variables (part I)

Fall 2024

Department of Statistics, North Carolina State University.

Random variables

Definition
Let Ω be a sample space and P a probability measure on Ω. A random variable (r.v.) X is a real-valued function from Ω to $\mathbb{R}$.

Example We throw a fair coin 4 times. A possible sample space Ω is the set of all $2^4 = 16$ possible sequences of H and T, i.e.,

$$\Omega = \{HHHH, HHHT, HHTH, \ldots, TTTT\}$$

Given Ω we can define many random variables X such as

- number of heads
- length of the longest run of heads
- length of the longest run of heads or tails
- location of the last head

Example Consider throwing a fair coin 100 times. A possible sample space Ω for this experiment is the set of all $2^{100} \approx 1.27 \times 10^{30}$ possible sequences of H and T of length 100. The following random variables,

- number of heads
- length of the longest run of heads
- length of the longest run of heads or tails
- location of the last head

only takes on integer values ranging from 0 to 100.

Example An election is coming up and you are tasked to poll 100 people on which of the two candidates A and B do they prefer. A possible sample space Ω for this experiment is the set of all 2^{100} possible sequences of A and B of length 100. Let X be the random variable mapping every sample point ω to the number of people who prefer candidate A. The range of this X is also identical to the range of the random variable X' for the number of heads in 100 throws of a fair coin.

Definition

Let Ω be a sample space and P a probability measure on Ω. A discrete random variable X is a real-valued function from Ω to some finite or countable $S \subset \mathbb{R}$. X also defines a probability measure P_X on the set of all subsets of S.

Example We throw a fair coin 4 times and let Ω be the set of all $2^4 = 16$ possible sequences of H and T of length 4. Let X be the random variable for the number of heads. Then

$$X \in \{0, 1, 2, 3, 4\}; \quad P_X(\{0\}) = P(\{TTTT\}) = 1/16$$
$$P_X(\{1\}) = P(\{TTTH, TTHT, THTT, HTTT\}) = 4/16,$$
$$P_X(\{2\}) = 6/16, \quad P_X(\{3\}) = 4/16, \quad P_X(\{4\}) = 1/16$$

Similarly, let Y be the random variable for the length of the longest run of heads. Then

$$Y \in \{0, 1, 2, 3, 4\}; \quad P_Y(\{0\}) = P(\{TTTT\}) = 1/16$$
$$P_Y(\{1\}) = P(\{TTTH, TTHT, THTT, HTTT, \dots\}) = 7/16,$$
$$P_Y(\{2\}) = P(\{HHTT, HTHH, TTHH, THHT, HHTH\}) = 5/16$$
$$P_Y(\{3\}) = 2/16, \quad P_Y(\{4\}) = 1/16$$

Note The above examples hopefully convince you that

- Given a sample space Ω, we can define many/infinite number of random variables.
- We can define, using quite different sample spaces Ω and Ω', the "same" random variable X.
- The range of values for X can be much much much smaller than the number of sample points in Ω.
- A random variable X also induce a corresponding probability measure.

Even for the same sample space Ω and probability measure P on Ω, different random variables (with identical range) induce different probability measures.

The above points imply the following important observation.

Observation If all you care about is the r.v. X, then the choice of the sample space Ω is generally not important (at least most of the time, and at least for this class). The only thing that matters is the probability measure induced by X.

pmf and cdf

Definition
Let X be a discrete random variable taking values in some finite or countable $S \subset \mathbb{R}$. X is associated with a probability mass function (pmf) $p\colon S \mapsto [0,1]$ satisfying

$$\sum_{x \in S} p(x) = 1.$$

X is also associated with a cumulative distribution function (cdf) $F\colon \mathbb{R} \mapsto [0,1]$ satisfying

$$F(x) = \sum_{\substack{z \in S \\ z \leq x}} p(x).$$

Note We interpret $p(x)$ and $F(x)$ as

$$p(x) = P(X = x) \quad \text{for } x \in S; \quad F(x) = P(X \le x) \quad \text{for } x \in \mathbb{R}$$

F is non-decreasing, i.e., $F(x) \le F(y)$ for $x \le y$ and

$$\lim_{x \to -\infty} F(x) = 0; \quad \lim_{x \to +\infty} F(x) = 1.$$

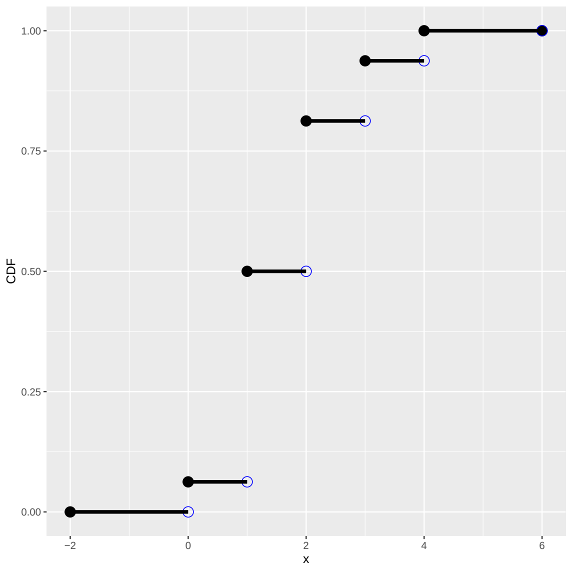
Example Recall the previous example of throwing a fair coin 4 times. Let Y be the random variable for the length of the longest run of heads. We saw the pmf f of Y earlier, i.e., with $f(y) = P(Y = y)$,

$$f(0) = \tfrac{1}{16},\ f(1) = \tfrac{7}{16},\ f(2) = \tfrac{5}{16},\ f(3) = \tfrac{2}{16},\ f(4) = \tfrac{1}{16}.$$

The cumulative distribution function for Y is (note the discontinuity at $y \in \{0, 1, 2, 3, 4\}$).

$$F(y) = P(Y \leq y) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{1}{16} & \text{if } 0 \leq y < 1 \\ \frac{8}{16} & \text{if } 1 \leq y < 2 \\ \frac{13}{16} & \text{if } 2 \leq y < 3 \\ \frac{15}{16} & \text{if } 3 \leq y < 4 \\ 1 & \text{if } y \geq 4 \end{cases}$$

The plot of the cdf for Y is given below.



Definition
Let X be a discrete random variable taking values in $S \subset \mathbb{R}$. Denote by p_X the pmf for X. Then for any function $g\colon \mathbb{R} \mapsto \mathbb{R}$, $Y = g(X)$ is also a discrete random variable taking values in $T = \{g(x)\colon x \in S\}$ with pmf p_Y defined by

$$p_Y(y) = P(Y = y) = \sum_{\substack{x \in S \\ g(x) = y}} p_X(x).$$

Example Let X be the random variable with pmf

$$p_X(0) = \frac{1}{16},\ p_X(1) = \frac{7}{16},\ p_X(2) = \frac{5}{16},\ p_X(3) = \frac{2}{16},\ p_X(4) = \frac{1}{16}.$$

Consider the r.v. $Z = (Y-2)^2$. Then

$$Z = 0 \iff Y = 2;\ Z = 1 \iff Y \in \{1,3\};\ Z = 4 \iff Y \in \{0,4\}.$$

Z thus has pmf p_Z given by

$$\begin{aligned} p_Z(0) &= p_X(2) = \frac{5}{16}, \\ p_Z(1) &= p_X(1) + p_X(3) = \frac{10}{16}, \\ p_Z(4) &= p_X(0) + p_X(4) = \frac{2}{16}. \end{aligned}$$

Expected values of random variables

Definition
Let X be a discrete random variable taking values in $S \subset \mathbb{R}$ with pmf p. Then for any function $g\colon \mathbb{R} \mapsto \mathbb{R}$, the mean or expected value of $g(X)$ is defined as

$$\mathbb{E}[g(X)] = \sum_{x \in S} g(x)p(x) = \sum_{x \in S} g(x)P(X = x).$$

Note When $g(x) = x$, the above expression simplifies to

$$\mathbb{E}[X] = \sum_{x \in S} xp(x) = \sum_{x \in S} xP(X = x).$$

The quantity $\mathbb{E}[X]$ is known as the mean of X.

Example Let Y be a random variable with pmf

$$P(Y = 1) = 1/16; \; P(Y = 2) = 7/16; \; P(Y = 3) = 5/16;$$
$$P(Y = 4) = 2/16; \; P(Y = 5) = 1/16$$

Evaluate $\mathbb{E}[Y]$ and $\mathbb{E}[1/Y]$.

Example Let Y be a random variable with pmf

$$P(Y=1)=1/16;\;\; P(Y=2)=7/16;\;\; P(Y=3)=5/16;$$
$$P(Y=4)=2/16;\;\; P(Y=5)=1/16$$

Evaluate $\mathbb{E}[Y]$ and $\mathbb{E}[1/Y]$.

$$\mathbb{E}[Y]=\sum_{y=1}^{5} yP(Y=y)=1/16+14/16+15/16+8/16+5/16\approx 2.69$$

$$\mathbb{E}[1/Y]=\sum_{y=1}^{5} \frac{1}{y}P(Y=y)=1/16+7/32+5/48+2/64+1/80\approx 0.43$$

We note that $1/\mathbb{E}[Y]\approx 0.372<\mathbb{E}[1/Y]$. This is always true, i.e., for any Y,

$$\mathbb{E}\left[\frac{1}{Y}\right]\geq\frac{1}{\mathbb{E}[Y]} \qquad \text{provided these quantities exist.}$$

In general, it is rarely (if ever) the case that $\mathbb{E}[g(X)] = g(\mathbb{E}[X])$.

The following is an important case where $\mathbb{E}[g(X)] = g(\mathbb{E}[X])$.

Proposition
Let X be a random variable and $g(X) = aX + b$ for some constants a and b. Then $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

A r.v. with infinite expectation

Let X be a discrete random variable with pmf

$$p(x) = P(X = x) = \frac{1}{x(x+1)} \quad \text{for } x = 1, 2, 3, \ldots,$$

A r.v. with infinite expectation

Let X be a discrete random variable with pmf

$$p(x) = P(X = x) = \frac{1}{x(x+1)} \quad \text{for } x = 1, 2, 3, \ldots,$$

We note that f is a valid pmf as it is non-negative and

$$\sum_{x=1}^{\infty} \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \Big(\frac{1}{x} - \frac{1}{x+1}\Big) = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \cdots = 1$$

A r.v. with infinite expectation

Let X be a discrete random variable with pmf

$$p(x) = P(X = x) = \frac{1}{x(x+1)} \quad \text{for } x = 1, 2, 3, \ldots,$$

We note that f is a valid pmf as it is non-negative and

$$\sum_{x=1}^{\infty} \frac{1}{x(x+1)} = \sum_{x=1}^{\infty}\Big(\frac{1}{x} - \frac{1}{x+1}\Big) = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \cdots = 1$$

However, $\mathbb{E}[X]$ does not exists as

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x p(x) = \sum_{x=1}^{\infty} \frac{1}{x+1}$$

which diverges to ∞. In general for a random variable X $\mathbb{E}[g(X)]$ is well-defined whenever

$$\sum_{x \in S} |g(x)| p(x) < \infty.$$

Definition
Let X be a discrete random variable with mean $\mu = \mathbb{E}[X]$. Assume $\mathbb{E}[X^2]$ exists. Then the variance of X is defined as

$$\sigma_X^2 = \mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mu^2.$$

The standard deviation of X is defined as $\sigma_X = \sqrt{\mathrm{Var}[X]}$.

Note As will be made abundantly clear in the subsequent part of this course (and in ST502), the mean and variance are the two most important summary statistics of any random variable. The mean and standard deviation is usually interpreted as summarizing the ”location/centrality” and ”spread” of a random variable, respectively.

Note The mean and the variance exist for almost all random variables of interest to us (in this class). There are random variables for which the mean and variance do not exist. In general the mean exists whenever the variance exists.

Note The equation $\mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mu^2$ implies the useful fact that $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$. Compare this with $\mathbb{E}[1/X] \geq 1/\mathbb{E}[X]$.

Example Let X be the random variable for the face that shows up when we roll a fair die. Then X has pmf

$$p(x) = \frac{1}{6} \quad \text{for } x = 1, 2, \ldots, 6.$$

We then have

$$\mathbb{E}[X] = \sum_{x=1}^{6} x p(x) = \frac{1}{6} \sum_{x=1}^{6} x = \frac{7}{2},$$

$$\mathbb{E}[X^2] = \sum_{x=1}^{6} x^2 p(x) = \frac{1}{6} \sum_{x=1}^{6} x^2 = \frac{91}{6},$$

$$\mathrm{Var}[X] = \sum_{x=1}^{6} (x - \mathbb{E}[X])^2 \times f(x) = \frac{1}{6} \sum_{x=1}^{6} \left(x - \tfrac{7}{2}\right)^2 = \frac{35}{12} = \frac{91}{6} - \left(\frac{7}{2}\right)^2.$$

37. The n candidates for a job have been ranked $1, 2, 3, \ldots, n$. Let X = the rank of a randomly selected candidate, so that X has pmf

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

(this is called the *discrete uniform distribution*). Compute $E(X)$ and $V(X)$ using the shortcut formula. [*Hint:* The sum of the first n positive integers is $n(n + 1)/2$, whereas the sum of their squares is $n(n + 1)(2n + 1)/6$.]

37. The n candidates for a job have been ranked $1, 2, 3, \ldots, n$. Let $X =$ the rank of a randomly selected candidate, so that X has pmf

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

(this is called the *discrete uniform distribution*). Compute $E(X)$ and $V(X)$ using the shortcut formula. [*Hint:* The sum of the first n positive integers is $n(n+1)/2$, whereas the sum of their squares is $n(n+1)(2n+1)/6$.]

$$\mathbb{E}[X] = \sum xp(x) = \sum_{x=1}^{n} \frac{x}{n} = \frac{1}{n} \sum_{x=1}^{n} x = \frac{n+1}{2},$$

$$\mathbb{E}[X^2] = \sum_{x=1}^{n} \frac{x^2}{n} = \frac{(n+1)(2n+1)}{6},$$

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(n-1)(n+1)}{12}.$$

Indicator random variables

Definition
Let Ω be a sample space and $A \subset \Omega$ be any event. The indicator random variable I_A for A is defined as

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}.$$

Very Important Let Ω be a sample space and P a probability measure on Ω. Then

$$P(A) = \mathbb{E}[I_A] \quad \text{for all events } A \subset \Omega.$$

In other words, expectation and probabilities are equivalent.

Expectation is a linear operation

Proposition

Let X be a discrete random variables with pmf p. Let $g_1, g_2, \ldots, g_m$ be real-valued functions on $\mathbb{R}$. Then

$$\begin{aligned}
\mathbb{E}[g_1(X) + \cdots + g_m(X)] &= \sum_{x \in S} (g_1(x) + \cdots + g_m(x)) p(x) \\
&= \sum_{x \in S} (g_1(x)p(x) + \cdots + g_m(x)p(x)) \\
&= \sum_{x \in S} g_1(x)p(x) + \cdots + \sum_{x \in S} g_m(x)p(x)) \\
&= \mathbb{E}[g_1(X)] + \cdots + \mathbb{E}[g_m(X)].
\end{aligned}$$

The previous result is a special cases of the following general result (which also holds for random variables more general than discrete random variables).

Proposition
Let $X_1, X_2, \ldots, X_m$ be discrete random variables. Let $g_1, g_2, \ldots, g_m$ be real-valued functions on $\mathbb{R}$. Then

$$\mathbb{E}[g_1(X_1) + \cdots + g_m(X_m)] = \mathbb{E}[g_1(X_1)] + \cdots + \mathbb{E}[g_m(X_m)].$$

As we will see many times throughout this class, expectations are generally "easy" to handle due to the above properties.

Example We now verify the expression

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

in the definition of $\mathrm{Var}[X]$. Indeed,

$$\begin{aligned}
\mathbb{E}[(X-\mu)^2] &= \mathbb{E}[X^2 - 2X\mu + \mu^2] \\
&= \mathbb{E}[X^2] - \mathbb{E}[2X\mu] + \mathbb{E}[\mu^2] \\
&= \mathbb{E}[X^2] - 2\mu \times \mathbb{E}[X] + \mu^2 \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 = \mathbb{E}[X^2] - \mu^2.
\end{aligned}$$

Variance of a linear function

Q. Let X be a random variable with variance σ_X^2. Let $Y = aX + b$ for some constants a and b. Find $\mathrm{Var}[Y]$.

Variance of a linear function

Q. Let X be a random variable with variance σ_X^2. Let $Y = aX + b$ for some constants a and b. Find $\mathrm{Var}[Y]$.

Denote $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. Then $\mu_Y = a\mu_X + b$.

Using the definition of the variance, we have

$$\mathrm{Var}[Y] = \mathbb{E}[(Y - \mu_Y)^2] = \mathbb{E}[(aX - a\mu_X)^2] = a^2\mathbb{E}[(X - \mu_X)^2] = a^2\sigma_X^2$$

Minimizing mean square error

Let X be a random variable. As X is random, we are interested in finding a fixed quantity c that is "most predictive" of X.

More specifically we want to find a constant c minimizing $\mathbb{E}[(X-c)^2]$.

Proposition
Let $h(c) = \mathbb{E}[(X-c)^2]$. *Then* $h(c)$ *is minimized at* $c^* = \mathbb{E}[X]$.

Indeed, let $\mu_X = \mathbb{E}[X]$. Then for any c

$$\begin{aligned}
\mathbb{E}[(X-c)^2] &= \mathbb{E}[(X-\mu_X+\mu_X-c)^2] \\
&= \mathbb{E}[(X-\mu_X)^2 + 2(X-\mu_X)(\mu_X-c) + (\mu_X-c)^2] \\
&= \mathbb{E}[(X-\mu_X)^2] + 2(\mu_X-c)\mathbb{E}[X-\mu_X] + (\mu_X-c)^2 \\
&= \mathbb{E}[(X-\mu_X)^2] + (\mu_X-c)^2 \\
&\geq \mathbb{E}[(X-\mu_X)^2] = \mathrm{Var}[X]
\end{aligned}$$

Example: Sum of m die

Let X be the random variable for the sum of the two rolls when a fair die is rolled twice. Then $X = X_1 + X_2$ where X_1 and X_2 are the value on the first and second roll, respectively. Note that

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \sum_{k=1}^{6} k \times P(X = k) = \sum_{k=1}^{6} k \times \frac{1}{6} = \frac{7 \times 6}{2 \times 6} = 3.5$$

and hence $\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 7$.

Example: Sum of m die

Let X be the random variable for the sum of the two rolls when a fair die is rolled twice. Then $X = X_1 + X_2$ where X_1 and X_2 are the value on the first and second roll, respectively. Note that

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \sum_{k=1}^{6} k \times P(X = k) = \sum_{k=1}^{6} k \times \frac{1}{6} = \frac{7 \times 6}{2 \times 6} = 3.5$$

and hence $\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 7$.

Suppose a fair die is now rolled m times and let X be the sum of these rolls. In a similar fashion to the above, let X_i be the value on the ith roll. Then

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_1 + X_2 + \cdots + X_m] \\ &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_m] = m \times \mathbb{E}[X_1] = \frac{7m}{2}.\end{aligned}$$

Example: Expectation of max vs max of expectation

Let X be the random variable for the number of heads in 4 toss of a fair coin. Then $Y = 4 - X$ is the random variable for the number of tails. The probability mass function for X and Y are identical, i.e.,

$$p_X(0) = \tfrac{1}{16},\ p_X(1) = \tfrac{4}{16},\ p_X(2) = \tfrac{6}{16},\ p_X(3) = \tfrac{4}{16},\ p_X(4) = \tfrac{1}{16}$$
$$p_Y(0) = p_X(4) = \tfrac{1}{16},\ p_Y(1) = p_X(3) = \tfrac{4}{16}, \ldots$$

We therefore have $\mathbb{E}[X] = \mathbb{E}[Y] = 2$.

Q. Now let $Z = \max\{X, Y\}$. What is $\mathbb{E}[Z]$?

The pmf for Z is now

$$p_Z(0) = 0,\ p_Z(1) = 0,\ p_Z(2) = \tfrac{6}{16},\ p_Z(3) = \tfrac{8}{16},\ p_Z(4) = \tfrac{2}{16}$$

and hence $\mathbb{E}[Z] = 44/16 = 2.75 > \max\{\mathbb{E}[X], \mathbb{E}[Y]\}$.

In summary, even if X and Y have the same expectation or mean, the larger of X and Y does not have that same expectation.

Example: paradoxical expectations

Q. We throw a fair four-sided die two times. Let X and Y be the random variables for the face showing on the first and second throw, respectively. Let $Z = X/Y$. Find $\mathbb{E}[Z]$.

Example: paradoxical expectations

Q. We throw a fair four-sided die two times. Let X and Y be the random variables for the face showing on the first and second throw, respectively. Let $Z = X/Y$. Find $\mathbb{E}[Z]$.

Let Ω be the sample space of all $4^2 = 16$ possibilities for (X, Y). These 16 sample points are equiprobable.

The table for values of Z is as follows

X vs Y	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$
$X = 1$	1	1/2	1/3	1/4
$X = 2$	2	1	2/3	1/2
$X = 3$	3	3/2	1	3/4
$X = 4$	4	2	4/3	1

Using the previous table,

$$\mathbb{E}[Z] = \frac{1}{16}(1 + 1/2 + 1/3 + \cdots + 1) = 1.3$$

Q. Since $Z = X/Y$ and $\mathbb{E}[Z] > 1$, we say that, on average, the value of the first throw is larger than the value of the second throw. Now swapping the label of X and Y in the above table yield the same table (and the same expected value), and so, on average, the value of the second throw is larger than the value of the first throw. Is there a contradiction ? Why or why not ?

Example: St Petersburg paradox

One interpretation of the expected value of a random variable is in terms of its utility function. For example, a gambler who plays a game of chances will want to maximize his/her expected winings.

Consider the following game. A fair coin is flipped repeatedly until it comes up heads the first time, after which the player wins 2^x dollars where x is the number of times the coin was flipped. The cost to enter the game is T dollars.

You can only play the game once.

Q1. Is say $T = 1000$ a fair price to pay to play this game ?

A. Suppose we pay T dollars to play. Then if $x = 1$, we win $2 - T$ dollars, if $x = 2$ we win $4 - T$ dollars, if $x = 3$ we win $8 - T$ dollars. We thus expect to win

$$\sum_{x=1}^{\infty}(2^x - T)p(x) = \sum_{x=1}^{\infty}(2^x - T)2^{-x} = \sum_{x=1}^{\infty}(1 - T2^{-x}) = \infty \text{ dollars.}$$

We can thus expect to be infinitely rich after playing this game, no matter how big T is.

Q2. What is the probability of losing money when $T = 1000$?

A. We will lose money if $2^x - T < 0$, i.e., if $x \leq 9$, i.e., we lose money with probability

$$p(1) + p(2) + \cdots + p(9) = 2^{-1} + 2^{-2} + \cdots + 2^{-9} = 1 - 2^{-9} = 0.998.$$

Example: Waiting for success

Peter and the Wolf play a simple game of dice as follows. Peter keeps throwing the fair die until he obtains the sequence 11 (two 1's in a row), after which he stop. Let X be the random variable for the number of throws before Peter stops.

Next is the Wolf. The Wolf also throws the fair die like Peter die, but now stops the first time the sequence 12 (a 1 followed by a 2) appears. Let Y be the number of throws before the Wolf stops. If $Y < X$ then the Wolf wins and he/she eats Peter. If $X < Y$ then Peter wins and they listen to Prokofiev.

Q. Who is "expected" to win ?

First consider Peter. Let X_0 be the expected number of throws to get two consecutive 1 (when the previous throw is not a 1), and let X_1 be the expected number of throws to get two consecutive 1 (when the previous throw was a 1). Then

$$X_0 = \tfrac{5}{6}(1 + X_0) + \tfrac{1}{6}(1 + X_1)$$
$$X_1 = \tfrac{5}{6}(1 + X_0) + \tfrac{1}{6} \times 1$$

This is a system of two equations in two unknowns. Hence

$$\begin{bmatrix} X_0 = 6 + X_1 \\ 5X_0 = 6X_1 - 6 \end{bmatrix} \Longrightarrow 36 - X_1 = 0 \Longrightarrow X_0 = 42.$$

Thus, on average, Peter will need to throw the die 42 times before seeing two consecutive 1.

Now consider the Wolf. Let X_0 be the expected number of throws to get a 1 followed by a 2 when we start the game. Let X_1 be the expected number of throws to get a 2 (when the previous throw is a 1). We then have

$$X_0 = \tfrac{5}{6}(1 + X_0) + \tfrac{1}{6}(1 + X_1)$$
$$X_1 = \tfrac{4}{6}(1 + X_0) + \tfrac{1}{6}(1 + X_1) + \tfrac{1}{6} \times 1.$$

This is once again a system of two equations in two unknown. Hence

$$\begin{bmatrix} X_0 = 6 + X_1 \\ 4X_0 = 5X_1 - 6 \end{bmatrix} \Longrightarrow 30 - X_1 = 0 \Longrightarrow X_1 = 36.$$

Thus, on average, the Wolf will only need to throw the die 36 times before seeing a 1 followed directly by a 2. Hopefully Peter is a good tree climber.

Example: Matching problem

Q. Suppose n people throw their hats in the air and the wind brings each of them back one hat at random. What is the expected number of people who received their hat back ?

Example: Matching problem

Q. Suppose n people throw their hats in the air and the wind brings each of them back one hat at random. What is the expected number of people who received their hat back ?

A. Let X be the random variable for the number of people who get his/her hat back. Let A_i be the event that the ith person received his/her hat back. Let I_{A_j} denote the indicator random variable for event A_j. Then

$$X = I_{A_1} + I_{A_2} + \cdots + I_{A_n}$$

We therefore have

$$\mathbb{E}[X] = \mathbb{E}[I_{A_1} + I_{A_2} + \cdots + I_{A_n}] = \mathbb{E}[I_{A_1}] + \cdots + \mathbb{E}[I_{A_n}].$$

Furthermore, for any j, $\mathbb{E}[I_{A_j}] = P(A_j) = \frac{1}{n}$. We therefore have $\mathbb{E}[X] = 1$, a value that does not depends on n.

Example: Gambler's ruin (simplified)

Q. Suppose a coin with probability p for heads in a single toss is tossed n times. How many times can we expect to see $k \geq 2$ heads in a row ?

Example: Gambler's ruin (simplified)

Q. Suppose a coin with probability p for heads in a single toss is tossed n times. How many times can we expect to see $k \geq 2$ heads in a row ?

A. Let X denote the random variable for the number of times k heads appear consecutively. Let A_j for $j = 1, 2, \ldots, n-k+1$ denote the event that the $j, j+1, \ldots, j+k-1$ toss all result in heads. Then $X = I_{A_1} + I_{A_2} + \cdots + I_{A_{n-k+1}}$. Furthermore

$$\mathbb{E}[I_{A_j}] = P(A_j) = p^k.$$

We therefore have

$$\mathbb{E}[X] = \sum_{j=1}^{n-k+1} \mathbb{E}[I_{A_j}] = (n-k+1) \times p^k.$$

As an example, let $n = 1040$, $k = 10$ and $p = 0.5$. Then we would expect to see $1031 \times 2^{-10} \approx 1.006 > 1$ time of a sequence of 10 heads in a row. In general, for a given n, we can expect to see at least one run of $k \geq \log(n/p)$ heads in a row.