Random Variables

A function $\mathrm{Z} : \Omega \to \mathbf{R}$ is called a random variable.

Two Dice

$\mathrm{Z}(x_1, x_2) = x_1 + x_2$.
$p_k = \mathbf{P}(\mathrm{Z} = k) = \mathbf{P}(\{\omega : \mathrm{Z}(\omega) = k\})$.

k	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Coloured Balls

$\Omega = \{k$ indistinguishable balls, n colours$\}$.
Uniform distribution.
$\mathrm{Z} =$ no. colours used.

$$p_m = \frac{\binom{n}{m}\binom{k-1}{m-1}}{\binom{n+k-1}{k}}.$$

If $k = 10, n = 5$ then

$p_1 = \frac{5}{1001}, p_2 = \frac{90}{1001}, p_3 = \frac{360}{1001}, p_4 = \frac{420}{1001}$,

$p_5 = \frac{126}{1001}$.

Binomial Random Variable $B_{n,p}$.

n coin tosses. $p = \mathbf{P}(Heads)$ for each toss.
$\Omega = \{H, T\}^n$.

$$\mathbf{P}(\omega) = p^k(1-p)^{n-k}$$

where k is the number of H's in ω.
$B_{n,p}(\omega) =$ no. of occurrences of H in ω.

$$\mathbf{P}(B_{n,p} = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

If $n = 8$ and $p = 1/3$ then

$p_0 = \frac{2^8}{3^8}$, $p_1 = 8 \times \frac{2^7}{3^8}$, $p_2 = 28 \times \frac{2^6}{3^8}$,

$p_3 = 56 \times \frac{2^5}{3^8}$, $p_4 = 140 \times \frac{2^4}{3^8}$, $p_5 = 56 \times \frac{2^3}{3^8}$,

$p_6 = 28 \times \frac{2^2}{3^8}$, $p_7 = 8 \times \frac{2}{3^8}$, $p_8 = \frac{1}{3^8}$

Poisson Random Variable $Po(\lambda)$.

$\Omega = \{0, 1, 2, \ldots, \}$ and

$$\mathbf{P}(Po(\lambda) = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \text{for all } k \geq 0.$$

This is a limiting case of $B_{n,\lambda/n}$ where $n \to \infty$.

$Po(\lambda)$ is the number of occurrences of an event which is individually rare, but has constant expectation in a large population.

Fix k, then

$$\begin{aligned}\lim_{n\to\infty}\mathbf{P}(B_{n,\lambda/n}=k) &= \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\ &= \frac{\lambda^k e^{-\lambda}}{k!}\end{aligned}$$

Explanation of $\binom{n}{k}\approx n^k/k!$ for fixed k.

$$\begin{aligned}\frac{n^k}{k!} &\geq \binom{n}{k}\\ &= \frac{n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\\ &\geq \frac{n^k}{k!}\left(1-\frac{k(k-1)}{2n}\right)\end{aligned}$$

Expectation (Average)

Z is a random variable. Its *expected value* is given by

$$\begin{aligned}\mathbf{E}(Z) &= \sum_{\omega\in\Omega} Z(\omega)\mathbf{P}(\omega)\\ &= \sum_k k\mathbf{P}(Z=k).\end{aligned}$$

Ex: **Two Dice**
$Z = x_1 + x_2$.

$$\mathbf{E}(Z) = 2\times\frac{1}{36}+3\times\frac{2}{36}+\cdots+12\times\frac{1}{36}=7.$$

Ex:10 indistinguishable balls, 5 colours. Z is the number of colours actually used.

$$\mathbf{E}(Z)=\frac{5}{1001}+2\times\frac{90}{1001}+3\times\frac{360}{1001}+4\times\frac{420}{1001}+5\times\frac{126}{1001}.$$

In general: n colours, m balls.
$Z_i = 1 \ \leftrightarrow$ colour i is used.
$Z = Z_1+\cdots+Z_n$= number of colours actually used.

$$\begin{aligned}\mathbf{E}(Z) &= \mathbf{E}(Z_1)+\cdots+\mathbf{E}(Z_n)\\ &= n\mathbf{E}(Z_1)\\ &= n\,\Pr(Z_1\neq 0)\\ &= n\left(1-\frac{\binom{n+m-2}{m}}{\binom{n+m-1}{m}}\right).\\ &= n\left(1-\frac{n-1}{n+m-1}\right)\\ &= \frac{mn}{n+m-1}.\end{aligned}$$

Alternative proof:

$$\begin{aligned}\mathbf{E}(Z) &= \sum_{k=1}^{n}\frac{k\binom{n}{k}\binom{m-1}{k-1}}{\binom{n+m-1}{m}}\\ &= n\sum_{k=1}^{n}\frac{\binom{n-1}{k-1}\binom{m-1}{k-1}}{\binom{n+m-1}{m}}\\ &= n\sum_{k-1=0}^{n-1}\frac{\binom{n-1}{k-1}\binom{m-1}{m-k}}{\binom{n+m-1}{m}}\\ &= \frac{n\binom{n+m-2}{m-1}}{\binom{n+m-1}{m}}\\ &= \frac{mn}{n+m-1}.\end{aligned}$$

Geometric

$\Omega = \{0, 1, 2, \ldots, \}$
$\mathbf{P}(k) = (1-p)^{k-1}p,\ \mathrm{Z}(k) = k.$

$$\begin{aligned}\mathbf{E}(\mathrm{Z}) &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\ &= \frac{p}{(1-(1-p))^2} \\ &= \frac{1}{p}\end{aligned}$$

= expected number of trials until success.

$$\left[\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}.\right]$$

Binomial $B_{n,p}$.

$$\begin{aligned}\mathbf{E}(B_{n,p}) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k} \\ &= np\sum_{k=1}^{n} \binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} \\ &= np(p+(1-p))^{n-1} \\ &= np.\end{aligned}$$

Poisson $Po(\lambda)$.

$$\begin{aligned}\mathbf{E}(Po(\lambda)) &= \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} \\ &= \lambda\sum_{k=1}^{\infty} \frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!} \\ &= \lambda.\end{aligned}$$

Suppose X, Y are random variables on the same probability space Ω.

Claim: $\mathbf{E}(\mathrm{X}+\mathrm{Y}) = \mathbf{E}(\mathrm{X}) + \mathbf{E}(\mathrm{Y})$.
Proof:

$$\begin{aligned}E(\mathrm{X}+\mathrm{Y}) &= \sum_{\alpha}\sum_{\beta}(\alpha+\beta)\mathbf{P}(\mathrm{X}=\alpha,\mathrm{Y}=\beta) \\ &= \sum_{\alpha}\sum_{\beta}\alpha\mathbf{P}(\mathrm{X}=\alpha,\mathrm{Y}=\beta) + \sum_{\alpha}\sum_{\beta}\beta\mathbf{P}(\mathrm{X}=\alpha,\mathrm{Y}=\beta) \\ &= \sum_{\alpha}\alpha\sum_{\beta}\mathbf{P}(\mathrm{X}=\alpha,\mathrm{Y}=\beta) + \sum_{\beta}\beta\sum_{\alpha}\mathbf{P}(\mathrm{X}=\alpha,\mathrm{Y}=\beta) \\ &= \sum_{\alpha}\alpha\mathbf{P}(\mathrm{X}=\alpha) + \sum_{\beta}\beta\mathbf{P}(\mathrm{Y}=\beta) \\ &= \mathbf{E}(\mathrm{X}) + \mathbf{E}(\mathrm{Y}).\end{aligned}$$

In general if $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ are random variables on Ω then

$$\mathbf{E}(\mathrm{X}_1+\mathrm{X}_2+\cdots+\mathrm{X}_n) = \mathbf{E}(\mathrm{X}_1)+\mathbf{E}(\mathrm{X}_2)+\cdots+\mathbf{E}(\mathrm{X}_n)$$

Binomial
Write $B_{n,p} = \mathrm{X}_1+\mathrm{X}_2+\cdots+\mathrm{X}_n$ where $\mathrm{X}_i = 1$ if the ith coin comes up heads.

$$\mathbf{E}(B_{n,p}) = \mathbf{E}(\mathrm{X}_1) + \mathbf{E}(\mathrm{X}_2) + \cdots + \mathbf{E}(\mathrm{X}_n) = np$$

since $\mathbf{E}(\mathrm{X}_i) = p \times 1 + (1-p) \times 0$.

Same probability space. $\mathrm{Z}(\omega)$ denotes the number of occurrences of the sequence H, T, H in ω.
$\mathrm{Z} = \mathrm{X}_1+\mathrm{X}_2+\cdots+\mathrm{X}_{n-2}$ where $\mathrm{X}_i = 1$ if coin tosses $i, i+1, i+2$ come up H, T, H respectively. So

$$\mathbf{E}(\mathrm{Z}) = \mathbf{E}(\mathrm{X}_1)+\mathbf{E}(\mathrm{X}_2)+\cdots+\mathbf{E}(\mathrm{X}_{n-2}) = (n-2)p^2(1-p),$$

since $\mathbf{P}(x_i = 1) = p^2(1-p)$.

m distinguishable balls, n boxes

$$\begin{aligned} Z &= \text{number of non-empty boxes.} \\ &= Z_1 + Z_2 + \cdots + Z_n \end{aligned}$$

where $Z_i = 1$ if box i is non-empty and $= 0$ otherwise. Hence,

$$\mathbf{E}(Z) = n\left(1 - \left(1 - \frac{1}{n}\right)^m\right),$$

since $\mathbf{E}(Z_i) = \mathbf{P}(\text{ box } i \text{ is non-empty}) = \left(1 - \left(1 - \frac{1}{n}\right)^m\right)$.

Random Walk: Suppose we do n steps of previously described random walk. Let Z_n denote the number of times the walk visits the origin. Then

$$Z_n = Y_0 + Y_1 + Y_2 + \cdots + Y_n$$

where $Y_i = 1$ if $X_i = 0$ – recall that X_i is the position of the particle after i moves.

But

$$\mathbf{E}(Y_i) = \begin{cases} 0 & i \text{ odd} \\ \binom{i}{i/2}2^{-i} & i \text{ even} \end{cases}$$

So

$$\begin{aligned} \mathbf{E}(Z_n) &= \sum_{\substack{0 \le m \le n \\ m \text{ even}}} \binom{m}{m/2} 2^{-m}. \\ &\approx \sum \sqrt{2/(\pi m)} \\ &\approx \frac{1}{2}\int_0^n \sqrt{2/(\pi x)}dx \\ &= \sqrt{2n/\pi} \end{aligned}$$

Consider the following program which computes the minimum of the n numbers $x_1, x_2, \ldots, x_n$.

begin
$min := \infty$;
for $i = 1$ **to** n **do**
begin
if $x_i < min$ **then** $min := x_i$
end
output min
end

If the x_i are all different and in random order, what is the expected number of times that that the statement $min := x_i$ is executed?

$\Omega = \{$permutations of $1, 2, \ldots, n\}$ – uniform distribution.
Let X be the number of executions of statement $min := x_i$.
Let

$$X_i = \begin{cases} 1 & \text{statement executed at } i. \\ 0 & \text{otherwise} \end{cases}$$

Then $X_i = 1$ iff $x_i = \min\{x_1, x_2, \ldots, x_i\}$ and so

$$\mathbf{P}(X_i = 1) = \frac{(i-1)!}{i!} = \frac{1}{i}.$$

[The number of permutations of $\{x_1, x_2, \ldots, x_i\}$ in which x_i is the largest is $(i-1)!$.] So

$$\begin{aligned} \mathbf{E}(X) &= \mathbf{E}\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \mathbf{E}(X_i) \\ &= \sum_{i=1}^n \frac{1}{i} \qquad (= H_n) \\ &\approx \log_e n. \end{aligned}$$

Independent Random Variables

Random variables X, Y defined on the same probability space are called independent if for all α, β the events $\{X = \alpha\}$ and $\{Y = \beta\}$ are independent.

Example: if $\Omega = \{0,1\}^n$ and the values of X, Y depend only on the values of the bits in disjoint sets Δ_X, Δ_Y then X, Y are independent.

E.g. if $X =$ number of 1's in first m bits and $Y =$ number of 1's in last $n - m$ bits.

The independence of X, Y follows directly from the disjointness of $\Delta_{\{X=\alpha\}}$ and $\Delta_{\{Y=\beta\}}$.

If $X = B_{n,p} =$ number of heads in n coin flips and $Y = n - B_{n,p}$ then X and Y are not independent. E.g. $\mathbf{P}(X = n) = p^n$ but $\mathbf{P}(X = n \mid Y = n) = 0$.

Now suppose the number of coin flips is the random variable $N = Po(\lambda)$. Let X be number of heads and Y be the number of tails. Let $q = 1 - p$.

$$\begin{aligned} \mathbf{P}(X = x, Y = y) &= \mathbf{P}(X = x, Y = y \mid N = x + y) \\ &\qquad \times \mathbf{P}(N = x + y) \\ &= \binom{x+y}{x} p^x q!^y \frac{\lambda^{x+y}}{(x+y)!} e^{-\lambda} \\ &= \frac{(\lambda p)^x (\lambda q)^y}{x! y!} e^{-\lambda}. \end{aligned}$$

$$\begin{aligned} \mathbf{P}(X = x) &= \sum_{n \geq x} \mathbf{P}(X = x \mid N = n) \mathbf{P}(N = n) \\ &= \sum_{n \geq x} \binom{n}{x} p^x q^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \frac{(\lambda p)^x}{x!} e^{-\lambda} \sum_{n-x \geq 0} \frac{(\lambda q)^{n-x}}{(n-x)!} \\ &= \frac{(\lambda p)^x}{x!} e^{-\lambda} e^{\lambda q} \\ &= \frac{(\lambda p)^x}{x!} e^{-\lambda p}. \end{aligned}$$

Similarly,

$$\mathbf{P}(Y = y) = \frac{(\lambda q)^y}{y!} e^{-\lambda q}$$

and so

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x) \mathbf{P}(Y = y)$$

for all x, y and the two random variables are independent!