Let us look at the Population Mean and Variance of the Bernoulli(θ) RV.

Let $X \sim$ Bernoulli(θ). Then,

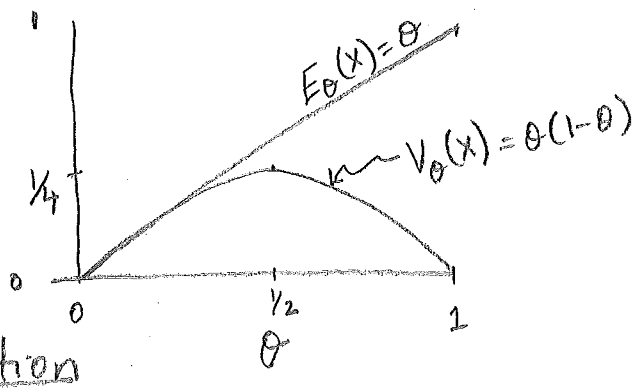
$$E(X) = \sum_{x=0}^{1} x f(x;\theta) = (0 \times (1-\theta)) + (1 \times (\theta)) = 0 + \theta = \theta$$

$$E(X^2) = \sum_{x=0}^{1} x^2 f(x;\theta) = (0^2 \times (1-\theta)) + (1^2 \times (\theta)) = 0 + \theta = \theta$$

$$V(X) = E(X^2) - (E(X))^2 = \theta - \theta^2 = \theta(1-\theta)$$

We subscript E and V by θ to emphasize parameter-specificity:

$$E_\theta(X) = \theta \qquad \text{and} \qquad V_\theta(X) = \theta(1-\theta).$$



Population Mean and Variance of the Uniform(0,1) RV.

Let $X \sim$ Uniform(0,1). Then,

$$E(X) = \int x \, dF(x) = \int_0^1 x f(x) dx = \int_0^1 x \cdot 1 \, dx = \frac{1}{2}\left(x^2\right]_0^1 = \frac{1}{2}(1-0) = \frac{1}{2}$$

$$E(X^2) = \int x^2 dF(x) = \int_0^1 x^2 f(x) dx = \int_0^1 x^2 \cdot 1 \, dx = \frac{1}{3}\left(x^3\right]_0^1 = \frac{1}{3}(1-0) = \frac{1}{3}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{4}{12} - \frac{3}{12} = \frac{4-3}{12} = \frac{1}{12}$$

Proposition [Winnings on Average]

Let $Y = r(X)$. Then,

$$E(Y) = E(r(X)) = \int r(x)\, dF(x).$$

" Think of playing a game where we draw $x \sim X$ and then I pay you $y = r(x)$. Then your average income is $r(x)$ times the chance that $X = x$, summed (or integrated) over all values of x.

Corollary [Probability is an Expectation]

Let A be an event and let $r(x) = \mathbb{1}_A(x)$. Recall that $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ if $x \notin A$. Then,

$$E(\mathbb{1}_A(X)) = \int \mathbb{1}_A(x)\, dF(x) = \begin{cases} \int_A f(x)\, dx = P(X \in A) = P(A) & \text{, if } X \text{ is contin. RV.} \\ \sum_{x \in A} f(x) = P(X \in A) = P(A) & \text{, if } X \text{ is discrete RV.} \end{cases}$$

Thus, probability is a special case of expectation. Now, recall our long-term relative frequency (LTRF) motivation for the axiomatic definition of probability and make the connection.

Exercise: Work through the "optional" material in Lab 3 and make the formal connections with the lectures.

* Fruit-bowl RVs and Probability of their realizations
* IID coin tossing experiments. (observe how independence and identity conspire to assign probabilities to events like 'HHTT' etc)

Now, we will focus on the NZ lotto & IID coin tosses to get a concrete introduction to Data, Dataspace, Statistics, Empirical Mass and Distribution functions.

NZ Lotto — Models for RVs, Data, Statistics.

Model 3 de Moivre $(1/k, 1/k, \ldots, 1/k)$ RV X has a discrete uniform distribution over $\{1, 2, \ldots, k\}$.

["Think of rolling a polygonal cylindrical die with k rectangular faces marked $1, 2, \ldots, k$" or "the first ball out of a Lotto machine with k balls numbered $1, 2, \ldots, k$".

Thus, PDF of X is:

$$f(x; (1/k, 1/k, \ldots, 1/k)) = \begin{cases} 0 & \text{if } x \notin \{1, 2, \ldots, k\} \\ 1/k & \text{if } x \in \{1, 2, \ldots, k\}. \end{cases}$$

Expectations:

Using the Faulhaber's formula for $\sum_{i=1}^{k} i^m$ with $m=1$ or $m=2$,

$$E(X) = \frac{1}{k}(1+2+\cdots+k) = \frac{1}{\cancel{k}} \frac{\cancel{k}(k+1)}{2} = \frac{k+1}{2}$$

$$E(X^2) = \frac{1}{k}(1^2+2^2+\cdots+k^2) = \frac{1}{k} \frac{k(k+1)(2k+1)}{6} = \frac{2k^2+3k+1}{6}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{2k^2+3k+1}{6} - \underbrace{\left(\frac{k+1}{2}\right)^2}$$

$$= \frac{2k^2+3k+1}{6} - \frac{k^2+2k+1}{4}$$

$$= \frac{8k^2+12k+4-6k^2-12k-6}{24}$$

$$= \frac{2k^2-2}{24} = \frac{k^2-1}{12}$$

Ex For NZ Lotto we model the first ball that pops out using de Moivre $(\frac{1}{40}, \frac{1}{40}, \ldots, \frac{1}{40})$ RV X. So, $P(X=1) = P(X=2) = \cdots = P(X=40)$

Probability, Data and Statistics

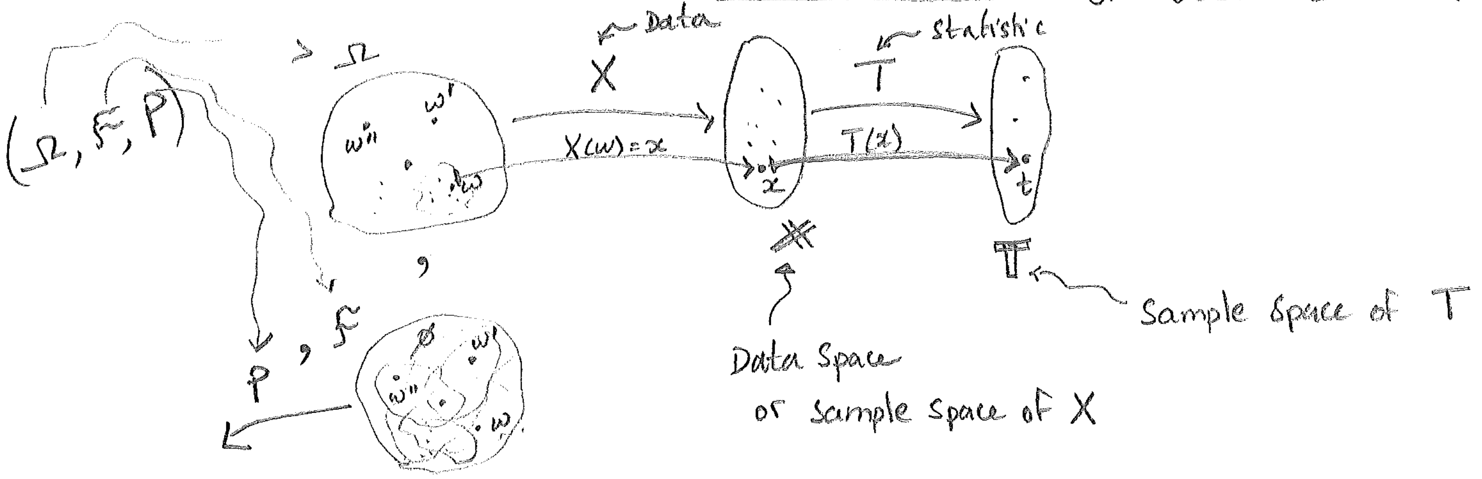
Defns.: Given some prob. triple $(\Omega, \mathcal{F}, P)$, let the function X "measure" the outcome ω from the sample space Ω.

$$X(\omega) : \Omega \longrightarrow \mathbb{X}$$

Often the measurements made by X are real numbers or vectors of real numbers (only cases seen in this course), i.e. $\vec{X} = (X_1, X_2, \ldots, X_n)$, such that X is a random variable (RV) or random vector $(R\vec{V})$.

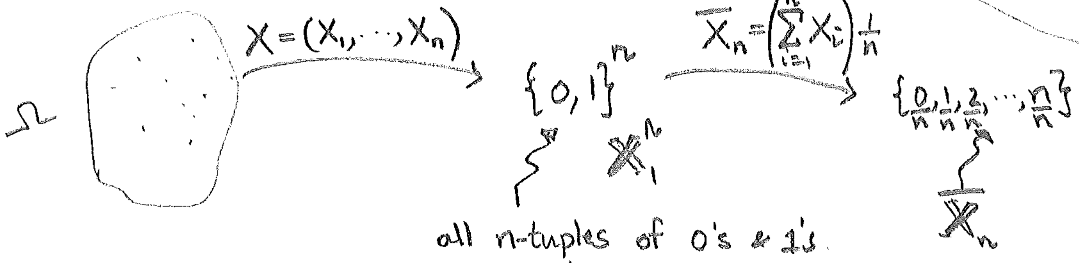
* Technical note: more generally X is a "measurable Map".

If X gives the finest empirical resolution of interest to the experimenter, then X is called data, $\mathbb{X}$ is called the data space (sample space of the data $\mathbb{X}$), $X(\omega) = x$ is the outcome ω of experiment measured by X and is called the observed data or realization of X, and finally a statistic $T(x) : \mathbb{X} \longrightarrow \mathbb{T}$ is a function (RV or $R\vec{V}$) of data X with $T(x) = t$ as the observed statistic of observed data x.



PRODUCT EXPERIMENTS

n IID Bernoulli($\theta=\frac{1}{2}$) RVs $X_1, X_2, \dots, X_n \overset{IID}{\sim}$ Bernoulli($\frac{1}{2}$) as $R\vec{V}$ $X=(X_1,\dots,X_n)$.

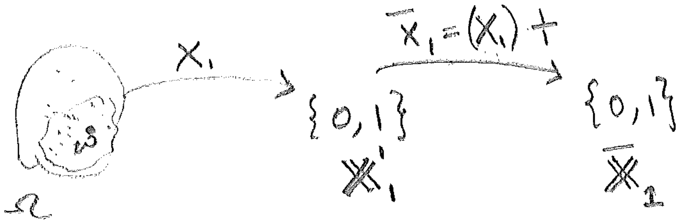


all n-tuples of 0's & 1's

$|\{0,1\}^n| = 2^n$

Ex :

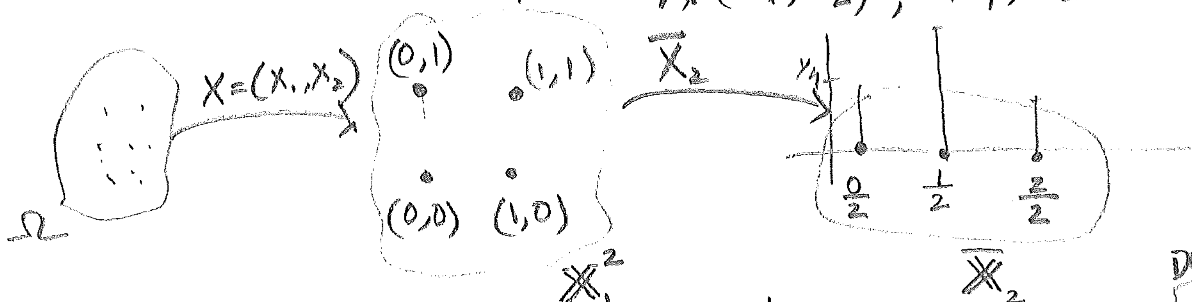
1 Bernoulli($\theta=\frac{1}{2}$) RV X_1



Suppose we observe tails where $(\Omega, \mathcal{F}, P)$ corresponds to a coin toss. then observed data $= X_1(\omega) = x = 0$ and obsered sample mean $\overline{X}_1(x) = \overline{x}_1 = 0$

Ex :

2 IID Bernoulli ($\theta=\frac{1}{2}$) $R\vec{V}$ $\mathbb{X}=(X_1,X_2)$, $X_1, X_2 \overset{IID}{\sim}$ Bernoulli($\frac{1}{2}$)



Due to independence and identity

$$P((X_1,X_2)=(0,0)) = P(X_1=0)\,P(X_2=0) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

Similarly, $P((X_1,X_2)=(0,1)) = P(X_1=0)\,P(X_2=1) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$

and $P((X_1,X_2)=(1,0)) = P((X_1,X_2)=(1,1)) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$

Don't forget Dfn of Prob:

? $P\{\omega : X_1(\omega)=0\} = 1-\theta = 1-\frac{1}{2} = \frac{1}{2}$ since $\theta=\frac{1}{2}$ here.

Now,

$$P(\overline{X}_2 = 0) = P\left(\{x=(x_1,x_2) \in \mathbb{X}_1^2 : \overline{X}_2(x) = \frac{x_1+x_2}{2} = 0\}\right)$$
$$= P(\{(0,0)\}) = \frac{1}{4}$$

///ly, $P(\overline{X}_2 = 1) = P(\{(1,1)\}) = \frac{1}{4}$

Prob of disjoint events

$$P(\overline{X}_2 = \frac{1}{2}) = P(\{(0,1),(1,0)\}) = P(\{(0,1)\}) + P(\{(1,0)\}) = \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$$

Ex : 3 IID Bernoulli($\frac{1}{2}$) $R\vec{V}$ ... etc ...

n IID deMoivre $(1/k, 1/k, \dots, 1/k)$ RVs

can be used to model the number on the first ball drawn in n Lotto trials. Lotto balls have numbers in $\{1,2,\dots,40\}$

$$X_1, X_2, \dots, X_n \overset{IID}{\sim} \text{de Moivre}(\underbrace{\tfrac{1}{40}, \tfrac{1}{40}, \dots, \tfrac{1}{40}}_{40 \text{ terms}})$$

That is $P(X_i = x_i) = \begin{cases} \frac{1}{40} & \text{if } x_i \in \{1,2,\dots,40\} \\ 0 & \text{if } x_i \notin \{1,2,\dots,40\} \end{cases}$

for each trial i, where $i \in \{1,2,\dots,n\}$. total number of trials.

In our NZ Lotto data, $n = 1114$.

Data Space denoted $\mathbb{X}_1^n = \{1,2,\dots,40\}^n$

Thus, due to independence and identity.

$$P((X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n))$$

$$= P(X_1 = x_1)\, P(X_2 = x_2) \cdots P(X_n = x_n)$$

$$= \begin{cases} \underbrace{\frac{1}{40} \cdot \frac{1}{40} \cdots \frac{1}{40}}_{n \text{ terms}} = \left(\frac{1}{40}\right)^n, & \text{if } (x_1, x_2, \dots, x_n) \in \mathbb{X}_1^n = \{1,2,\dots,40\}^n \\ 0, & \text{if } (x_1, x_2, \dots, x_n) \notin \mathbb{X}_1^n = \{1,2,\dots,40\}^n \end{cases}$$

Thus, every observation in our data space has equal prob. $\left(\frac{1}{40}\right)^n$

Our observed data $x = (\underbrace{4, 3, 11, 35, \dots, 16, 6, 8, 35}_{n = 1114 \text{ terms}})$ for NZ Lotto data up to now.

Note that our observed data x is just <u>one point</u> in our data space $\mathbb{X}_1^{1114} = \{1,2,\dots,40\}^{1114}$ and there are lots of other possible realizations of our data.

There are exactly 40^{1114}

Exercise: Is data a statistic?

In other words, is the RV X for which the realization is the observed data $X(\omega) = x$, a statistic?

[Hint: consider the identity map, $T(x) = x : \mathbb{X} \to \mathbb{T} = \mathbb{X}$.]

Some Elementary Statistics

Sample Mean

From a given sequence of RVs $X_1, X_2, \cdots, X_n$, we may obtain another RV called the n-samples mean or simply the sample mean:

$$T_n((X_1, X_2, \cdots, X_n)) = \overline{X}_n((X_1, X_2, \cdots, X_n)) := \frac{1}{n} \sum_{i=1}^{n} X_i$$

For brevity, we write

$$\overline{X}_n((X_1, X_2, \cdots, X_n)) \text{ as } \overline{X}_n$$

and its realisation

$$\overline{X}_n((x_1, x_2, \cdots, x_n)) \text{ as } \overline{x}_n .$$

By properties of Expectations we have seen before:

$$E(\overline{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) .$$

Furthermore, if every X_i in $X_1, X_2, \cdots, X_n$ is identically distributed with the same expectation, say $E(X_1)$, then:

$$E(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} n \cdot E(X_1) = E(X_1) .$$

Similarly, we can show that:

$$V(\bar{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) \qquad \text{(recall properties of Variance!)}$$

Furthermore, if the original sequence of RVs $X_1, X_2, \ldots, X_n$ is independently distributed then:

$$V(\bar{X}_n) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) \qquad \text{(recall properties of variance)}.$$

Finally, if the sequence $X_1, X_2, \ldots, X_n$ is independently and identically distributed with the same variance, say $V(X_1)$, then:

$$V(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\, n\, V(X_1) = \frac{1}{n} V(X_1).$$

— x —

Sample Variance

Sample Standard Deviation = $\sqrt{\text{Sample Variance}}$

Order statistics.
EMF
EDF
Histograms

— x —