# All of Statistics

My proposed solutions to the book:

**All of Statistics - A Concise Course in Statistical Inference**.

I really needed a proper refresher on statistics.

Let me know if you find any mistakes! I am sure there are plenty. :(

## Contents

# 1 Chapter 3 - Expectation

## Exercises

### 3.1

Define $X$ as the wealth after $n$ games. The probability of winning and losing is the same for each outcome so $p = 1/2$.

$$\mathbb{E}[X] = \frac{1}{2} \cdot 2c + \frac{1}{2} \cdot \left(\frac{1}{2}\right) c = c + \frac{c}{4} = \frac{5}{4}c$$

We expect to have $5/4 \cdot c$ after $n$ games. We can also verify this result with a simulation in **R**.

```
> games = sample(c(2, 0.5), size = 1000000, replace = TRUE)
> mean(games)
[1] 1.250973
> 5/4
[1] 1.25
```

R

### 3.2

**Claim**. $\text{Var}(X) = 0$ if and only if $\mathbb{P}(X = c) = 1$ for some constant $c$.

Proof.
$\Rightarrow$) Set $c = \mu$ and assume $\text{Var}(X) = 0$, which means that

$$\mathbb{E}[(X - \mu)^2] = 0 \quad \Longrightarrow \quad \int (x - \mu)^2 dF(x) = 0.$$

This can only be 0 when $x = \mu = c$ for the entire domain of $X$. Hence $\mathbb{P}(X = c) = 1$.

$\Leftarrow$) Assume $\mathbb{P}(X = c) = 1$. When calculating the expectation:

$$\mu = \mathbb{E}[X] = \int c dF(x) = c$$

When calculating the variance:

$$\text{Var}(X) = \int (x - \mu)^2 dF(x) = \int (c - c)^2 dF(x) = 0,$$

since $x = c$ for all $x$ in the domain of $X$. $\square$

**3.3**

Let $X_1, \dots, X_n \sim U(0,1)$ and define $Y = \max(X_1, \dots, X_n)$. We will calculate $\mathbb{E}[Y]$. It is not stated in the exercise, but we will assume that the $X_i$ are independent. Finding the CDF for $Y$.

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leqslant y) \\
&= \mathbb{P}(\max(X_1, \dots, X_n) \leqslant y) \\
&= \mathbb{P}(X_1 \leqslant y) \cap \ldots \cap \mathbb{P}(X_n \leqslant y) \\
&= \mathbb{P}(X_1 \leqslant y)\mathbb{P}(X_2 \leqslant y) \cdots \mathbb{P}(X_n \leqslant y) && \text{(Independence)} \\
&= (F_X(y))^n
\end{aligned}$$

Differentiating to get the PDF for $Y$.

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(F_X(y))^n = n(F_X(y))^{n-1} f_X(y)$$

Since $X_i$ are uniformly distributed, we know that $F_X(y) = y$ and $f_X(y) = 1$, so:

$$f_Y(y) = ny^{n-1}$$

Now we can calculate the expectation of $Y$.

$$\mathbb{E}[Y] = \int_0^1 y \cdot ny^{n-1} dy = n \int_0^1 y^n dy = n \Big[\frac{y^{n+1}}{n+1}\Big]_0^1 = \frac{n}{n+1}$$

Confirming this result with a numeric simulation in **R**.

```
> # 3.3
> N = 1000000
> U1 = runif(N)
> U2 = runif(N)
> U3 = runif(N)
> U4 = runif(N)
> U5 = runif(N)
> U6 = runif(N)
> U7 = runif(N)
> U8 = runif(N)
> U9 = runif(N)
> U10 = runif(N)
> Y = pmax(U1, U2, U3, U4, U5,
+          U6, U7, U8, U9, U10)
> mean(Y)
[1] 0.9091151
> # Theoretical Result
> 10/11
[1] 0.9090909
```

R

As we can see, the theoretical result is very close to the simulated result for $n = 10$.

**3.4** - Random Walk

A particle starts in the origin and jumps left, a step of -1, with probability $p$ and jumps right, a step of 1, with probability $1-p$. The expected location will be:

$$\mathbb{E}[X] = (-1)p + (1)(1-p) = -p + 1 - p = 1 - 2p$$

To calculate the variance, we start by finding the second moment:

$$\mathbb{E}[X^2] = (-1)^2 p + (1)^2(1-p) = p + 1 - p = 1$$

So the variance is:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - (1-2p)^2 = 1 - (1 - 4p + 4p^2) = 4p - 4p^2$$

**3.5**

Tossing a fair coin until we get H. Finding the expected number of tosses. The reasoning is as follows. We get H on the first toss with probability 1/2, first H on the second toss with probability $1/2^2 = 1/4$ and so on. The pattern becomes as follows, for the first 7 cases:

| **Tosses** | **Outcome** | **Probability** |
|---|---|---|
| 1 | $\{H\}$ | 1/2 |
| 2 | $\{TH\}$ | 1/4 |
| 3 | $\{TTH\}$ | 1/8 |
| 4 | $\{TTTH\}$ | 1/16 |
| 5 | $\{TTTTH\}$ | 1/32 |
| 6 | $\{TTTTTH\}$ | 1/64 |
| 7 | $\{TTTTTTH\}$ | 1/128 |

Define $T$ to be the number of tosses to get H.

$$\begin{aligned}\mathbb{E}[T] &= (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) + \ldots + (k)\left(\frac{1}{2^k}\right) + \ldots \\ &= \sum_{k=1}^{\infty} \frac{k}{2^k} \\ &= 2\end{aligned}$$

Not delving in to the mathematics of the infinite sum, but it can be shown that this sum becomes 2 which will be the expected number of tosses to get a H. Here is a numeric approximation in **R**.

```
> sumApprox = 0
> for (k in 1:1000) {
+   sumApprox = sumApprox + k/2^k
+ }
> sumApprox
[1] 2
```

R

**3.6 Theorem -** The Rule of the Lazy Statistician
Proving the following result for the discrete case. Let $Y = r(X)$, then

$$\mathbb{E}[Y] = \mathbb{E}[r(X)] = \sum_x r(x) f(x)$$

PROOF.