Statistics for Data Science
Unit 4 Homework: Random Variables
Student: Richard Hitchens

1. **Best Game in the Casino**

Information provided:
Let,
H = a coin flip that lands on heads
T = a coin flip that lands on tails

The coin is fair, so P(H) = P(T) = 0.5

The game involves 3 flips, so the possible results of the game are: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

Let N be a discrete random variable equal to the number of heads you get when playing the game. Based on the possible results of the game N can take the values: {0,1,2,3}

The winnings for N=0 is W_0 = \$0, N=1 is W_1 = \$2, N=2 is W_2 = \$4 and N=3 is W_3 = \$X, where X is unknown.

Expected winnings = E(W) = \$6.

(a)

Find W_3 = \$X

Can do so by back solving E(W) = sum(W_i * P(N=i)), for i = 0, 1, 2 ,3.

Based on the 8 possible results of the game we know that:
P(N=0) = 1/8, i.e. {TTT}
P(N=1) = 3/8, i.e. {HTT, THT, TTH}
P(N=2) = 3/8, i.e. {HHT, HTH, THH}
P(N=3) = 1/8, i.e. {HHH}

Inputting all information provided:

\$6 = \$0 * 1/8 + \$2 * 3/8 + \$4 * 3/8 + \$X * 1/8

Therefore,

X = (\$48/8 - \$6/8 - \$12/8) / (1/8) = \$30

(b)

Let F(W) = the cumulative probability function of W.

Find F(W)

Based on the probabilities of N in part (a):

$$F(W) = \begin{cases} 0 & for \quad W < 0 \\ 0.125 & for \quad 0 \leq W < 2 \\ 0.5 & for \quad 2 \leq W < 4 \\ 0.875 & for \quad 4 \leq W < 30 \\ 1 & for \quad W \geq 30 \end{cases}$$

2. Processing Pasta

Information provided:

L is a continuous random variable with the following probability density function:

$$f(l) = \begin{cases} 0, & l \leq 0 \\ l/2, & 0 < l \leq 2 \\ 0, & 2 < l \end{cases}$$

(a)

The cumulative probability function for *L* is defined as:

$$F(L) = \int_{-\infty}^{L} f(l)dl$$

but *f(l)* is only defined between 0 and 2 otherwise it is 0, so we need to find:

$$F(L) = \int_{0}^{L} f(l)dl, \quad 0 < L \leq 2$$

$$F(L) = \int_{0}^{L} l/2\, dl \quad = L^2/4$$

Therefore, the complete expression for *F(L)* is:

$$F(L) = \begin{cases} 0, & L \leq 0 \\ L^2/4, & 0 < L \leq 2 \\ 1, & 2 < L \end{cases}$$

(b)

Find *E(L)*

The expected length of pasta is defined as:

$$E(L) = \int l\, f(l)\, dl$$

Again just need to consider the range between 0 and 2.

$$E(L) = \int_0^2 l\, f(l)\, dl = \int_0^2 l^2/2 = (1/3) * 2^3/2 - 0 = 4/3$$

3. The Warranty is Worth It

Information provided:

The life span (*t* years) of a particular (shoddy) server is a continuous random variable, with a uniform probability distribution between 0 and 1 years, so define *T ~ U(0,1)*.

If the server lasts only *t* years, which is less than a year the manufacturer will pay you *x*:

$$x = 100(1\text{-t})^{1/2}$$

Let *X* be the random variable representing the payout from the contract.

(a)

From the information provided, the probability density function for *T* is:

$f(t) = 1$, for $0 \leq t \leq 1$ and 0 otherwise

so the cumulative distribution function of T is:

$F(t) = P(T \leq t) = t$, for $0 \leq t \leq 1$

The standard definition of conditional probability is:

Conditional probability = $P(B|A) = P(B \cap A) / P(A)$

where $P(B \cap A)$ = the probability the server survives from 6 months to 9 months
$P(A)$ = the probability the server survives 6 months

Therefore using *F(t)*

$P(B \cap A) = F(9/12) – F(6/12) = 9/12 – 6/12 = ¼$,

which is just the same as surviving 3 months, i.e. $F(3/12) = ¼$ because the distribution is uniform.

P(A) = F(6/12) = 1/2

So P(B|A) = (1/4) / (1/2) = 1/2

(b)

Find $F(x) = P(X \leq x)$

Because x is a function of t and we know distribution of t:

$P(X \leq x) = P(X \leq x(t)) = P(T \leq t)$

However, because $x(t)$ is negatively related to t

$P(X \leq x(t)) = P(T \geq t) = 1 - P(T \leq t)$

From (a)

$$F(T) = \begin{cases} 0, & T < 0 \\ T, & 0 \leq T < 1 \\ 1 & T \geq 1 \end{cases}$$

Transform from $x(t)$ to $t(x)$:

$$t = 1 - (x/100)^2$$

when $x=0, t=1$ and when $x=100, t=0$.

From above $F(X) = 1 - F(T)$

Therefore $F(X)$ is defined as

$$F(X) = \begin{cases} 0, & X < 0 \\ (X/100)^2, & 0 \leq X < 100 \\ 1, & X \geq 100 \end{cases}$$

(c)

Find $E(X)$

There are two main ways you can calculate $E(X)$:

(1)

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

(2)

$$E(X) = \int_{-\infty}^{\infty} x(t)\, f(t)\, dt$$

Approach (1)

From *F(X)* in (b), *f(x)* by differentiation is therefore defined as $2x/100^2$ between 0 and 100, so

$$E(X) = \int_0^{100} x * (2x/100^2)\, dx = \int_0^{100} 2x^2/100^2 dx = \left[2/3\, x^3/100^2\right] \begin{matrix} 100 \\ 0 \end{matrix}$$

E(X) = 66.66

Approach (2)

f(t) is defined on the range 0 to 1, so

$$E(X) = \int_0^1 100(1-t)^{1/2}.1\, dt = 100 \int_0^1 (1-t)^{1/2} dt$$

Let $u = 1 - t$, so $t = 1\text{-}u$ and $dt/du = -1$, so

$$E(X) = 100 \int_0^1 -u^{1/2}\, du = 100 * \left[-2/3(1-t)^{3/2}\right]_0^1 = 0 - (100 * -2/3 * 1)$$

E(X) = 66.66

i.e. both approaches produce the same answer.

4. The Baseline for Measuring Deviations

(a)

The value of *t* that will minimize *E(Y)* will be the *E(X)* because all the deviations of the random variable *X* are centered around this expectation, so the average deviation from a large sample will in theory head to zero in limit.

(b)

Accordingly, the value of *Y* in this instance in limit would be *0*.

5. Optional Advanced Exercise

Information provided:

X is a continuous random variable with probability density function *f(x)*
h(x) is an invertible function where h^{-1} is differentiable
Y = h(X) is a continuous random variable

so

y = h(x)
x = h^{-1}(y)

P(Y ≤ y) = P(Y ≤ h(x)) = P(X ≤ x)

$$P(X \le x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-h(\infty)}^{h(x)} f(h^{-1}(y))\,\frac{d(h^{-1}(y))}{dy} dy$$

which on the right hand side is now all a function of y, where

$$g(y) = f(h^{-1}(y))\,\frac{d(h^{-1}(y))}{dy}$$

So

$$P(Y \le y) = \int_{-\infty}^{x} f(x)\,dx = \int_{-h(\infty)}^{y} g(y)\,dy$$

Therefore *g(y)* is the probability density function of *Y*.