Expected Value and Variance of Random Variables

Stat 241

Example 1. Suppose a student has taken 10 courses and received 5 A's, 4 B's, and 1 C. Using the traditional numerical scale where an A is worth 4, a B is worth 3, and a C is worth 2, what is this student's GPA (grade point average)?

The key idea here is that the mean is a ____________ of ________________ times ______________________.

We can write this as

$$\text{mean} = \sum \text{value} \cdot \text{probability}$$

Expected Value

For a discrete random variable this translates to

$$\mathrm{E}(X) = \sum x f(x)$$

where the sum is taken over all possible values of X and f is the pmf for X.

The mean of a random variable also goes by another name: **expected value**. We can denote the mean of X by either μ_X or $\mathrm{E}(X)$.

Example 2. Let X be discrete random variable with probablilities given in the table below.

value of X	0	1	2
probability	0.2	0.5	0.3

What is the mean (expected value) of X?

Example 3. A local charity is holding a raffle. They are selling 1000 raffle tickets for \$5 each. The owners of five of the raffle tickets will win a prize. The five prizes are valued at \$25, \$50, \$100, \$1000, and \$2000. Let X be the value of the prize associated with a random raffle ticket (\$0 for non-winning tickets).

a. Waht is the probability of winning a prize?

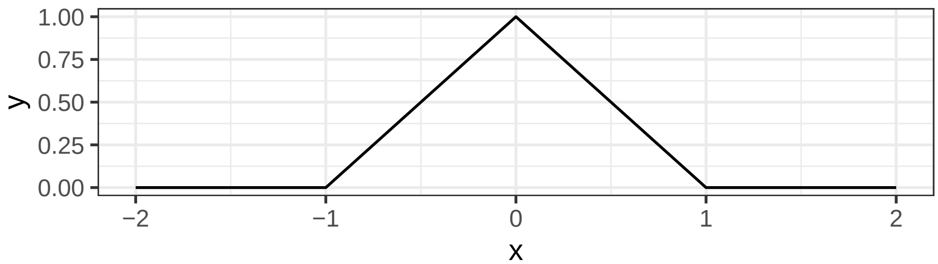
b. What is the prabability of winning the grand prize?

c. What is the expected value of the prize?

When working with a continuous random variable, we replace the sum with an integral and replace the probabilities with our density function to get the following definition:

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

Example 4. the mean of a symmetric triangle distribution on $[-1, 1]$.



$$\begin{aligned}
\mathrm{E}(X) &= \int_{-1}^{1} x f(x)\, dx \\
&= \int_{-1}^{0} x(x-1)\, dx + \int_{0}^{1} x(1-x)\, dx \\
&= \int_{-1}^{0} x^2 - x)\, dx + \int_{0}^{1} x - x^2)\, dx \\
&= \frac{x^3}{3} - \frac{x^2}{2}\Big|_{-1}^{0} + \frac{x^2}{2} - \frac{x^3}{3}\Big|_{0}^{1} \\
&= \frac{1}{3} - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = 0
\end{aligned}$$

We could also calculate this numerically in R:

```
library(mosaicCalc)
f <- makeFun( (1 - abs(x)) * (abs(x) <= 1) ~ x)
xf <- makeFun( x * f(x) ~ x )
integrate(xf, -1, 1)
```

```
## 0 with absolute error < 3.7e-15
```

```
F <- antiD( x * f(x) ~ x, lower.bound = -1)
F(-1)   # should be 0
```

```
## [1] 0
```

```
F(1)    # should be the expected value -- also 0.
```

```
## [1] 0
```

Variance

Arguing similarly, we can compute the variance of a discrete or continuous random variable using

- discrete: $\mathrm{Var}(X) = \sigma_X^2 = \sum_x (x - \mu_X)^2 f(x)\; dx$

*continuous: $\mathrm{Var}(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\; dx$

These can be combined into a single definition by writing

$$\mathrm{Var}(X) = \mathrm{E}((X - \mu_X)^2)\ .$$

Note: It is possible that the sum or integral used to define the mean (or the variance) will fail to converge. In that case, we say that the random variable has no mean (or variance) or that the mean (or variance) fails to exist.[1]

Example 5. Compute the variance of symmetric triangle distribution on $[-1, 1]$.

```
f <- makeFun( (1 - abs(x)) * (abs(x) <= 1) ~ x)
xxf <- makeFun( (x-0)^2 * f(x) ~ x )
integrate(xxf, -1, 1)
```

```
## 0.1666667 with absolute error < 1.9e-15
```

```
G <- antiD( (x-0)^2 * f(x) ~ x)
G(1) - G(-1)
```

```
## [1] 0.1666667
```

[1] Actually, we will require that $\int_{\infty}^{\infty} |x| f(x)\; dx$ converges and $\int_{\infty}^{\infty} |x|^2 f(x)\; dx$ converges. If these integrals (or the corresponding sums for discrete random variables) fail to converge, we will say that the distribution has no mean (or variance).

Useful identity

Some simple algebraic manipulations of the sum or integral above shows that

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \tag{1}$$

Example 6. Compute the mean and variance of the random variable with pdf given by

$$g(x) = \frac{3x^2}{8} [\![\, x \in [0,2] \,]\!] \,.$$

```
g <- makeFun( (3 * x^2/8 ) * (0 <= x & x <= 2) ~ x )
m <- antiD( x * g(x) ~ x, lower.bound = 0)(2)  # all in one step instead of defining F or G
m
```

```
## [1] 1.5
```

```
v <- antiD( (x - m)^2 * g(x) ~ x, m = m, lower.bound = 0)(2)
v
```

```
## [1] 0.15
```

```
# here's the alternate computation
antiD( x^2 * g(x) ~ x, lower.bound = 0)(2) - m^2
```

```
## [1] 0.15
```

As with data, the standard deviation is the square root of the variance.

More Practice

1. You are invited to play the following game. You draw two chips without replacement from a jar containing 5 red, 5 blue, and 5 green chips. If both chips have the same color you win \$5. If the two chips have different colors, you win \$3. On average, how much will you win per game? It costs you \$4 to play the game. On average, would you win money, lose money, or break even playing this game?

2. Toss a fair coin 3 times. Let X be the number of heads produced. Find the pmf for X and use it to find the average number of heads produced when the coin is tossed 3 times.

3. Repeat (3) above under the assumption that the coin is biased and only has a 1/4 chance of producing a head.

4. Let $f(x) = 2e^{-2x}$ on $(0, \infty)$. Create this function in R and use R to integrate it on $[0, \infty)$. Note that e^x is expressed in R as `exp(x)`.

a. Is $f(x)$ is a pdf?

b. Use R to compute $\text{P}(-1 \leq X \leq 3)$.

c. Use R to compute the cumulative distribution function F and use F to compute $\text{P}(3 \leq X \leq 10)$.

d. Use R to compute the mean of X.