Topic 17: Expectation and Moments

02-680: Essentials of Mathematics and Statistics

November 14, 2024

What is a "Statistic"?

Definition: Anything that can be computed from the collected data
(i.e., must be observable).

Statistics deals with data.

Goal: Make inferences based on data

This process can be divided into three overlapping phases

(1) **Collecting data** — collect data in experiments; Preceded by forming hypotheses about phenomena of interest

(2) **Describing data** — describe the results

(3) **Analyzing data** — infer from the results the strength of the evidence with respect to the hypotheses

Generally two types of statistics

Point statistic — a single value computed from data (e.g., $\overline{X_n}$)

Interval or range statistics — an interval $a \leq x \leq b$ computed from the data

Note that a statistic is itself a random variable because a new experiment will produce new data to compute it.

1 Expectation

Given a probability distribution $p(X)$ of a random variable X, how can we summarize the distribution with a single value?

The expectation of a random variable is a number that attempts to capture the center of $p(X)$ and can be interpreted as the long-run average of many independent samples from the given distribution.

For a (discrete) random variable X, the ***expected value*** is

$$\mathbb{E}[X] = \sum_{x \in range(X)} x \cdot p(X = x).$$

We sometimes also call this the ***mean*** or ***first moment*** of the variable.

For instance, in our previous examples, for a pair of 4-sided die:

$$\mathbb{E}[C] = 2 \cdot \frac{1}{16} + 3 \cdot \frac{2}{16} + 4 \cdot \frac{3}{16} + 5 \cdot \frac{4}{16} + 6 \cdot \frac{3}{16} + 7 \cdot \frac{2}{16} + 8 \cdot \frac{1}{16} = \frac{80}{16} = 5$$

Note that in this case the expected value happens to be one of the outcomes, but if we consider a random variable D which is the outcome of a single 6-sided die:

$$\mathbb{E}[D] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

which is not a possible outcome.

For a continuous random variable, again we swap a sum for an integral:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

So if we look at random variable $U \sim Uniform(-1, 3)$, and thus $f(u) = \frac{1}{4}$.

$$\mathbb{E}[U] = \int_{-\infty}^{\infty} x \cdot \frac{1}{4} dx = \frac{1}{8} x^2 \Big|_{-1}^{3} = \frac{9}{8} - \frac{1}{8} = 1$$

2 Variance and Standard Deviation

In addition to the mean, many times we want to know something about how "far" the likely outcomes are from the mean, in the normal distribution from last week this visually is about how narrow the bell is. To do this we define the ***variance*** of a random variable as

$$\mathbb{V}[X] = \mathbb{E}\left[[X - \mathbb{E}[X]]^2 \right]$$

The variance is the expected squared difference between a random variables and its mean.

The ***standard deviation*** of a variable is the square root of the variance:

$$\sigma(X) = \sqrt{\mathbb{V}[X]} = \sqrt{\mathbb{E}\left[[X - \mathbb{E}[X]]^2 \right]}$$

If we look at a first coin, F:

$$\mathbb{E}[F] = \frac{1}{2}$$

The difference between X and $\mathbb{E}[F]$ is $+\frac{1}{2}$ and $-\frac{1}{2}$ (with equal probability).

$$\mathbb{V}[F] = \mathbb{E}\left[[F - \mathbb{E}[F]]^2\right] = \left(\frac{1}{2}\left(+\frac{1}{2}\right)^2 + \frac{1}{2}\left(-\frac{1}{2}\right)^2\right) = \frac{1}{4}$$

and thus

$$\sigma[F] = \sqrt{\mathbb{V}[F]} = \sqrt{\frac{1}{4}} = \frac{1}{2}.$$

What if instead we tried to simulate a coin flip into variable S. But we made a mistake and now we get 1 or 0 with probability 0.4995, but in some cases we output 100 (with probability 0.001). What are $\mathbb{E}[S], \mathbb{V}[S], \sigma[S]$?

$$\mathbb{E}[S] = 0 \cdot 0.4995 + 1 \cdot 0.4995 + 100 * 0.001 = 0 + 0.4995 + 0.1 = 0.5995$$

$$\mathbb{V}[S] = \mathbb{E}\left[[S - \mathbb{E}[S]]^2\right]$$

$$= \left(0.4995 \cdot 0.5995^2\right) + \left(0.4995 \cdot 0.4005^2\right) + \left(0.01 \cdot 99.4005^2\right) \approx 0.180 + 0.080 + 9.880 \approx 10.14$$

$$\sigma[S] = \sqrt{\mathbb{V}[S]} \approx \sqrt{10.14} \approx 3.184$$

So clearly the variance is much higher, but how often do we actually expect to see a value of 100? If $p(S = 100) = 0.001$ thats approximately one in every 1000 throws, if we throw much less than that we may think its a fair coin. That leads the question of how we measure these things in practice.

3 Properties of Expectation and Variance

3.1 Summation

For some set $X_1, X_2, ..., X_n$ of independent random variables

$$\mathbb{E}\left[\sum_{i\in[n]} X_i\right] = \sum_{i\in[n]} \mathbb{E}[X_i] \qquad \mathbb{V}\left[\sum_{i\in[n]} X_i\right] = \sum_{i\in[n]} \mathbb{V}[X_i]$$

3.2 Linearity

For some random variable X and constants $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b \qquad \mathbb{V}[aX + b] = a^2\,\mathbb{V}[X]$$

4 Conditional Expectation and Variance

4.1 Discrete

$$\mathbb{E}\left[X \mid Y = y\right] = \sum_x \left(x \cdot p(X = x \mid Y = y)\right)$$

$$\mathbb{V}\left[X \mid Y = y\right] = \sum_x \Big(\left[x - \mathbb{E}\left[X \mid Y = y\right]\right]^2 \cdot p(X = x \mid Y = y)\Big)$$

4.2 Continuous

$$\mathbb{E}\left[X \mid Y = y\right] = \int_{-\infty}^{\infty} \left(x \cdot f(x \mid y) dx\right)$$

$$\mathbb{E}\left[X \mid Y = y\right] = \int_{-\infty}^{\infty} \Big(\left[x - \mathbb{E}\left[X \mid Y = y\right]\right]^2 \cdot f(x \mid y) dx\Big)$$

Example. Assume $X \sim Uniform(0,1), Y \mid X = x \sim Uniform(x,1)$.

$$\begin{aligned}
\mathbb{E}[Y \mid X = x] &= \int_{-\infty}^{\infty} \left(y \cdot f(y \mid x) dy\right) \\
&= \int_x^1 \left(y \cdot \frac{1}{1-x} dy\right) \\
&= \frac{1}{1-x} \int_x^1 \left(y dy\right) \\
&= \frac{1}{1-x} \cdot \left[\frac{1}{2} y^2\right]_x^1 \\
&= \frac{1}{1-x} \cdot \left[\frac{1}{2} - \frac{1}{2} x^2\right] \\
&= \frac{1 - x^2}{2(1-x)} = \frac{(1+x)}{2}
\end{aligned}$$

$$
\begin{aligned}
\mathbb{V}[Y \mid X = x] &= \int_{-\infty}^{\infty} \Big([y - \mathbb{E}\left[Y \mid X = x\right]]^2 \cdot f(y \mid x) \Big)\, dy \\
&= \int_x^1 \left(\left[y - \frac{(1+x)}{2} \right]^2 \cdot \frac{1}{1-x} \right) dy \\
&= \frac{1}{1-x} \int_x^1 \left(\frac{(2y-(1+x))^2}{4} \right) dy \\
&= \frac{1}{1-x} \int_x^1 \left(\frac{\left(4y^2 - 4(1+x)y + (1+x)^2\right)}{4} \right) dy \\
&= \frac{1}{1-x} \int_x^1 \left(y^2 - (1+x)y + \frac{(1+x)^2}{4} \right) dy \\
&= \frac{1}{1-x} \left[\frac{y^3}{3} - \frac{(1+x)}{2} y^2 + \frac{(1+x)^2}{4} y \right]_x^1 \\
&= \frac{1}{1-x} \left[\frac{y^3}{3} + \frac{y(1+x)((1+x)-2y)}{4} \right]_x^1 \\
&= \frac{1}{1-x} \left[\frac{1-x^3}{3} + \frac{(1+x)^2(x-1))}{4} \right]_x^1 \\
&= \frac{1}{1-x} \left[\frac{(1-x)(1+x+x^2)}{3} - \frac{(1+x)^2\,(1-x)}{4} \right] \\
&= \frac{(1+x+x^2)}{3} - \frac{(1+x)^2}{4}
\end{aligned}
$$

5 Covariance & Correlation

Given two random variables X and Y with expectations $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$ respectively. The ***covariance*** between the two is defined as

$$Cov(X, Y) = \mathbb{E}\left[(X - \mu_x)(Y - \mu_Y)\right] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Assume also that the standard deviations are σ_X and σ_Y, then the ***correlation*** of X and Y is defined as

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

You can think of correlation as a normalized version of covariance.