Introducing the Concept of Expectation

Dmitri Nesteruk
QUANTITATIVE ANALYST

@dnesteruk

Goals:

Understand the notion of an *expectation* and associated statistical measures.

Overview

Expectation

Mean

Law of the Unconscious Statistician

Variance

Moments and the Moment Generating Function

Joint Distributions

Covariance and Correlation

Expectation

If you measure a newborn baby, how tall do you expect it to be?

We don't know, since it's a random variable

Measure n newborn babies (the more the better) and average out the result

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

This is the *mean* or *average* baby height

So, given a random variable X representing a baby's height, $\bar{x}$ is its *expected value* (51cm/20in, actually)

Expectation of a Random Variable

The expectation (expected value, mean, average) of a random variable is a weighted average of all possible values weighted by their respective probabilities.

Expectation Example

Suppose a stock price can increase by 1 point with $P_U = 0.6$, stay in place with $P_S = 0.1$ and go down by 1 point with $P_D = 0.3$

If X represents the change in stock price, what is $E[X]$?

$$E[X] = 1 \cdot 0.6 + 0 \cdot 0.1 + 0.3 \cdot (-1) = 0.3$$

In other words, on average, we'd expect the stock price to go up by 0.3 points

Mean

For a bounded discrete random variable X with probability function f, the expectation of X is

$$E[X] = \sum_x xf(x)$$

Also referred to as the *mean* of X or the *expected value* of X.

Mean of a Bernoulli Random Variable

A rv X has a Bernoulli distribution when X can only take values 0 or 1 with $P(X = 1) = p$

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p$$

Note that the expectation does not have to be a possible value of the variable

Existence of Mean

Given a variable X with a pf f

Consider

$\sum_{\text{Positive } x} xf(x)$ and $\sum_{\text{Negative } x} xf(x)$

If at least one of these is finite, then $E[X]$ is said to exist, defined as

$$\sum_{\text{All } x} xf(x)$$

If both of the sums are infinite, mean does not exist

Existence of Mean Example

Let X be a rv with pf

$$f(x) = \begin{cases} \frac{1}{2|x|(|x|+1)}, & x = \pm 1, \pm 2, \pm 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Check the sums

$$\sum_{-1}^{-\infty} x \frac{1}{2|x|(|x|+1)} = -\infty$$

$$\sum_{1}^{\infty} x \frac{1}{2|x|(|x|+1)} = \infty$$

Both sums are infinite $\therefore E[X]$ does not exist

Expectation for a Continuous Distribution

For a continuous bounded random variable, a sum turns into an integral

If X is a bounded continuous rv with pdf f,

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

also called the *expected value* or *mean*

Continuous Expectation Example

Suppose the time X until a part fails has the pdf

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Then

$$E[X] = \int_0^1 x \cdot 2x \, dx = \int_0^1 2x^2 \, dx = \frac{2}{3}$$

The expectation of the distribution with pdf f is 2/3

Existence of Continuous Mean

If either $\int_0^{\infty} xf(x)\,dx$ or $\int_{-\infty}^{0} xf(x)\,dx$ is finite,

The mean/expectation/expected value of X exists and is defined as

$$\int_{-\infty}^{\infty} xf(x)\,dx$$

If both integrals are infinite, then the mean does not exist

Existence of Continuous Mean Example

Consider the Cauchy distribution with pdf

$$f(x) = \frac{1}{\pi(1+x^2)}$$

This is a pdf, i.e. $\int_{-\infty}^{\infty} f(x) = 1$ **because**

$$\frac{d(\tan^{-1} x)}{dx} = \frac{1}{1+x^2}$$

Calculate the two integrals:

$\int_0^{\infty} \frac{x}{\pi(1+x^2)}\, dx = \infty$ **and** $\int_{-\infty}^{0} \frac{x}{\pi(1+x^2)}\, dx = -\infty$

Thus the mean does not exist

Function of a Random Variable

If X is a rv with pdf f, consider a real-valued function $r(X)$

To find the expected value of $r(X)$, we

- Let $Y = r(X)$
- Determine probability distribution of Y
- Determine $E[Y]$

$E[r(x)] = E[Y] = \int_{-\infty}^{\infty} yg(y)\,dy$
if the expectation exists

Expection of Function Example

Given a rv X with pdf of $f(x) = 3x^2, 0 < x < 1$

Let $r(x) = 1/x$

First we calculate pdf of $Y = r(X)$:

$$g(y) = 3y^{-4}, \qquad y > 1$$

This gives us the mean

$$E[Y] = \int_0^\infty y \cdot 3y^{-4}\, dy = \frac{3}{2}$$

These manipulations are unnecessary due to LOTUS

Law of the Unconscious Statistician

If X is a rv and $r(x)$ is a real-valued function of a real variable then, if X has a discrete distribution,

$$E[r(x)] = \sum_{All\ x} r(x)f(x)$$

If X has a continuous distribution,

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

Expectation of Function Example (LOTUS)

Given a rv X with pdf of $f(x) = 3x^2, 0 < x < 1$

Let $r(x) = 1/x$

$E[Y] = \int_0^1 r(x)f(x)\,dx = \int_0^1 \frac{1}{x} \cdot 3x^2\,dx = 3/2$

Note: in general, $E[f(X)] \neq f(E[X])$

Properties of Distributions

If X is a rv for which $E[X]$ exists,

If $X = c$ with probability 1, $E[X] = c$

If $Y = aX + b$ where a and b are finite constants (linear function),

$$E[Y] = aE[X] + b$$

If $X_1, \dots, X_n$ are n rvs s.t. $E[X_i]$ is finite, then

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$$
$$E[X_1 \times \dots \times X_n] = E[X_1] \times \dots \times E[X_n]$$

Variance

A mean does not completely describe a distribution

A uniform distribution from -1 to 1 and a standard normal distribution have same mean, but are drastically different

We need another measure of how ‘spread out’ the distribution is

Variance

If X is a rv with finite mean $\mu = E[X]$, the *variance* of X is defined as

$$V[X] = E[(X-\mu)^2]$$

If X has infinite mean or if $E[X]$ does not exist, $V[X]$ does not exist.

The *standard deviation* of X is the non-negative square root of $V[X]$ if the variance exists:

σ – standard deviation, σ^2 – variance

Variance Example

Consider a coin flip that yields a 0 or a 1 with equal probability $P = 0.5$

$\mu = E[X] = 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5$

Define $Y = (X - \mu)^2$**. Then** $V[X] = E[Y]$

Compute pf of Y**:**

x	0	1
y	0.25	0.25
$f(y)$	0.5	0.5

$V[X] = E[Y] = 0.5(0.25 + 0.25) = 0.25$

Standard deviation $\sigma = \sqrt{0.25} = 0.5$

Alternative Variance Calculation Method

We can expand variance calculation

$$\begin{aligned} V[X] &= E[(X-\mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

In other words,

$$V[X] = E[X^2] - (E[X])^2$$

Additive Property of Variance

Given independent random variables $X_1, \dots, X_n$ with finite means and constants $a_1, \dots, a_n$ and b,

$$V[a_1X_1 + \cdots + a_nX_n + b] = a_1^2V[X_1] + \cdots + a_n^2V[X_n]$$

Notice how b does not contribute to the variance

Moments

For a random variable X and $k \in \mathbb{N}$, the expectation $E[X^k]$ is called the kth moment of X

The mean $E[X]$ is the first moment of X

The kth moment exists iff $E[|X|^k] < \infty$

Central Moment

An nth central moment μ_n is the moment of a probability distribution about the mean

$\mu_n = E[(X - E[X])^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x)\, dx$

$\mu_0 = 1$ **(because** $\int 1 \cdot f(x)\, dx = 1$**)**

$\mu_1 = 0$ **(because** $E\left[X - E[X]\right] = E[X] - E\left[E[X]\right] = 0$**)**

μ_2 **is the variance**

Moment Generating Function

Given a rv X, for each $t \in \mathbb{R}$, the *moment generating function* (mgf) of X is defined as

$$\psi(t) = E[e^{tX}]$$

For $n \in \mathbb{N}$, the nth moment of X is the nth derivative of the mgf at $t = 0$, i.e.,

$$E[X^n] = \psi^{(n)}(0)$$

Calculating an MGF

Suppose X has a pdf $f(x) = e^{-x}$ for $x > 0$

Let's find the mgf:

$$\psi(t) = E[e^{tX}] = \int_0^\infty e^{tx} e^{-x}\, dx$$

$$= \int_0^\infty e^{(t-1)x}\, dx$$

The integral diverges for $t \geq 1$, so mgf is

$$\psi(t) = \frac{1}{1-t}$$

for $t < 1$

Calculating variance from MGF

Mgf is $\psi(t) = \frac{1}{1-t}$

Calculate derivatives:

$\psi'(t) = \frac{1}{(1-t)^2}$ **and** $\psi''(t) = \frac{2}{(1-t)^3}$

Armed with this, we can calculate variance:

$$\begin{aligned} E[X^2] - (E[X])^2 &= \psi''(0) - [\psi'(0)]^2 \\ &= 2 - 1 = 1 \end{aligned}$$

Thus, $V[X] = 1$

Means and Variances of Some Distributions

Distribution	Pmf	Mean	Variance
Uniform (disc)	$\frac{1}{b-a+1}$ for $x = a, \ldots, b$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometric	$(1-p)^k p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Hypergeometric	$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$n\frac{r}{N}$	$n\frac{r}{N}\frac{(N-r)}{N}\frac{N-n}{N-1}$

Means and Variances of Some Distributions

Distribution	Pdf	Mean	Variance
Uniform (cont)	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma	$\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Beta	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Joint Distributions

Interested in several random variables related to each other

E.g., if you pick a random family and measure their earnings and how many cars they have, those things *might* be related

Instead of one variable, we consider two or more

We'll focus on 2 rvs

Joint Distribution Example

Suppose we roll *two* dice

We want the probability that first dice is a 1 and second is a 3

$$P(X = 1, Y = 3) = \frac{1}{36}$$

A general calculation can be expressed as a joint pmf

Joint Probability Mass Function

Given two discrete random variables X and Y, their joint probability mass function (pmf) is defined as

$$P_{XY}(x, y) = P(X = x, Y = y)$$

The joint pmf need to satisfy

$$\sum_{x \in X, y \in Y} P_{XY}(x, y) = 1$$

Marginal PMF

The joint pmf has all the information regarding the distribution of X and Y

We can try to 'extract' the pmf of X from P_{XY}:

$$P_X(x) = P(X = x) = \sum_{y_i} P(X = x, Y = y_i)$$

We call $P_X(x)$ the *marginal* pmf of X.

Marginal PMF Example

Consider rvs X and Y with joint pmf

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	$1/6$	$1/4$	$1/8$
$X = 1$	$1/8$	$1/6$	$1/6$

Let's find $P_X(x)$

- $P_X(0) = \frac{1}{6} + \frac{1}{4} + \frac{1}{8} = 13/24$
- $P_X(1) = 1 - P_X(0) = 11/24$

$$P_X(x) = \begin{cases} \frac{13}{24}, & x = 0 \\ \frac{11}{24}, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Joint Cumulative Distribution Function

The joint cumulative distribution function of two random variables X and Y is defined as

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

We expect that

$$F_{XY}(\infty, \infty) = 1$$
$$F_{XY}(-\infty, y) = 0, \qquad \text{for any } y$$
$$F_{XY}(x, -\infty) = 0, \qquad \text{for any } x$$

Functions of 2 or More Random Variables

Sometimes we have a function which involves two or more rvs

- E.g., sum of two dice being rolled
- $r(X,Y)$

LOTUS for two discrete rvs

$$E[r(X,Y)] = \sum_{x_i, y_i} r(x_i, y_i) P_{XY}(x_i, y_i)$$

Joint Probability Density Function

Two random variables X and Y are *jointly continuous* if there exists a non-negative function f_{XY} s.t. for any set $A \in \mathbb{R}^2$,

$$P\big((X,Y) \in A\big) = \iint_A f_{XY}(x,y)\,dx\,dy$$

The function f_{XY} is called the *joint probability density function* of X and Y. Naturally,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$$

Joint PDF Example

Suppose X and Y have joint pdf

$$f_{XY}(x, y) = x + cy^2, \qquad 0 \le x \le 1, 0 \le y \le 1$$

Let's find the constant c

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = \int_0^1 \int_0^1 x + cy^2\, dx\, dy$$

$$= \left[\frac{y}{2} + \frac{cy^3}{3}\right]_{y=0}^{y=1} = \frac{1}{2} + \frac{1}{3}c = 1$$

Therefore, $c = 3/2$

Marginal PDFs

Similarly to marginal pmfs, marginal pdfs involve leaving one of the two variables alone, i.e.

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\,, \qquad \text{for all } x$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\,, \qquad \text{for all } y$$

This lets us find a pdf for one of the two (or more) variables from a joint pdf

Marginal PDF Example

Given the joint pdf

$$f(x, y) = 10x^2y, \qquad 0 \le y \le x \le 1$$

Let's find $f_X(x)$ and $f_Y(y)$

- Note the domains of x and y!

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = \int_0^x 10x^2y\, dy = 5x^4$$

$$f_Y(y) = \int_y^1 10x^2y\, dx = \frac{10}{3} y(1 - y^3)$$

Search for Related Variables

Does smoking cause cancer?

Need to take measurements of cancer and amounts smoked

- Also variables that influence the result

Do the two sets of values vary together (as one grows, so does another)?

- Calculate covariance

Can we get a standardized measure of how related they are?

- Calculate correlation

Reminder: correlation does not imply causation

Covariance

If X and Y are rvs with finite means $E[X] = \mu_X$ and $E[Y] = \mu_Y$, the covariance of X and Y is defined as

$$\text{cov}(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

if the expectation exists.

Covariance Example

Let X, Y have a joint pdf

$$f(x) = \begin{cases} 2xy + 0.5, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Let's compute $\operatorname{cov}(X, Y)$

First we need $\mu_X = \mu_Y$

$$\mu_X = \int_0^1 \int_0^1 (2x^2y + 0.5)\, dy\, dx = 7/12$$

Then,

$$\operatorname{cov}(X, Y)$$
$$= \int_0^1 \int_0^1 \left(x - \frac{7}{12}\right)\left(y - \frac{7}{12}\right)(2xy + 0.5)\, dy\, dx$$
$$= 1/144$$

What is Covariance?

Covariance gives a numerical measure to the degree to which X and Y vary together

The magnitude of $\text{cov}(X, Y)$ is influenced by X and Y

Correlation is a much better, standardized measure

Correlation

Given random variables X and Y with finite variances σ_X^2 and σ_Y^2, the *correlation* of X and Y is defined as

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

Correlation Values

Correlation values range from -1 to 1

$\rho = 0$ **means variables are uncorrelated (we expect this for independent variables)**

$\rho = 1$ **means as one variable grows/shrinks, so does another, perfectly**

- Correlation measures only a linear relationship

$\rho = -1$ **means variables have a perfect inverse relationship (as one grows, another shrinks)**

Summary

A random variable has an expected value $E[X]$ also called mean or average

Variance ($E[(X - \overline{x})^2]$) measures how 'spread out' the values are

Two or more variables can have a joint pmf/pdf/cdf

- Marginal functions can be extracted from joint function

Covariance/correlation are measures of a linear relationship between two variables