CS 107, Probability, Spring 2020
Lecture 35

Michael Poghosyan
mpoghosyan@aua.am

AUA

04 May 2020

Content

- Partial Numerical Characteristics: Variance and the SD of a r.v.
- Numerical Characteristics of Important Distributions

Variance of a r.v.

Reminder: Expectation

In our previous lecture we talked about the Expectation: Expectation was showing the Mean of our r.v. values. Say, if we have a r.v. X with the following PDF:

The Variance of a R.V.

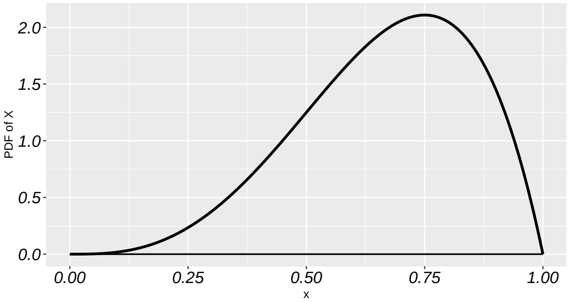


Figure: PDF $f(x) = 20 \cdot x^3(1-x), x \in [0,1]$

Then the range of possible values of X is $[0,1]$. But this doesn't mean that the Expected/Mean value of X is 0.5. This is because the chances are higher that we will get values close to 1, so the Mean, Expectation will be close to 1.

Reminder: Expectation

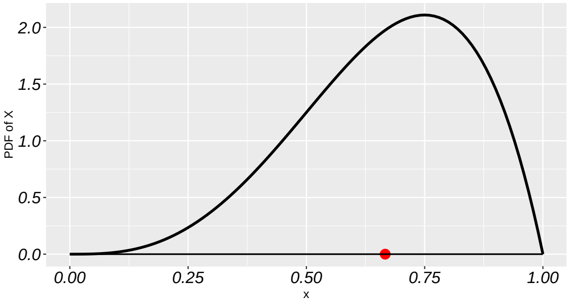


Figure: PDF $f(x) = 20 \cdot x^3(1-x), x \in [0,1]$, red dot is the Expected Value

Then the range of possible values of X is $[0,1]$. But this doesn't mean that the Expected/Mean value of X is 0.5. This is because the chances are higher that we will get values close to 1, so the Mean, Expectation will be close to 1.

Variance: Intro

Let us now consider the following r.v.s,

$$X \sim \mathcal{N}(0, 0.5^2), \quad Y \sim \mathcal{N}(0, 1), \quad Z \sim \mathcal{N}(0, 2^2),$$

then their Expectations will be 0,

$$\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = 0,$$

and only information about the Mean will not help us to differentiate between these distributions, although that distributions differ a lot:

Variance: Intro

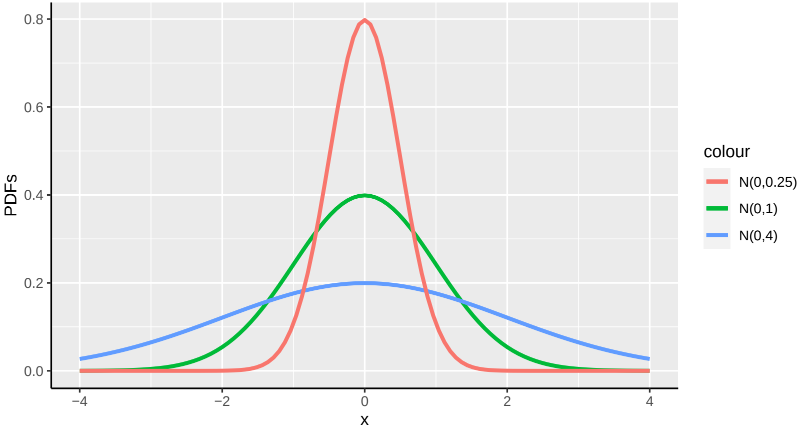


Figure: PDFs of different Normals with the same Mean.

Variance: Intro

Other examples: Assume that we have a contract to buy AMZN shares, an we know that the Expected price for AMZN stock at the end of the next week will be $\$2.5K$. This is, of course, an important information. And we imagine that the price at the end of the next week will be around $\$2.5K$. Is this giving us a nice picture what the actual price will be? No, of course.

Variance: Intro

Consider 2 cases:

- The AMZN stock price at the end of the next week can be anything from the interval $[1000, 4000]$, uniformly. In this case, the Expected price will be 2500. But we have no much information about the concrete price we will have at the end of the next week, because of the wide range of possibilities. So the risk is high.
- The AMZN stock price at the end of the next week can be anything from the interval $[2490, 2510]$, uniformly. In this case too, the Expected price will be 2500. But here the information is much more precise about the price, and the risk is low.

Variance: Intro

So knowing the Mean, Expected Value of the r.v. is an important thing, but this will not give sufficient information about the distribution, about the r.v.

Variance is the second most important numerical (partial) characteristics of the r.v., and it is measuring the dispersion or concentration, spread of values of the r.v. around its Mean.

Note: We will denote the Variance of a r.v. X by

$$Var(X) \qquad \text{or} \qquad \sigma_X^2,$$

and read it as the Variance or the Dispersion of X.

Deviations from the Mean

To define the Variance, we define first the Deviation from the Mean:

$$X - \mathbb{E}(X).$$

This is a r.v. showing how far the values of X are from the mean, and satisfying

$$\mathbb{E}\left(X - \mathbb{E}(X)\right) = 0.$$

Now, we want to give just one number to describe this r.v., to describe the Deviation from the mean. To that end, we square these Deviations, and calculate the Mean of the Squared Deviations:

Variance of a r.v.

The Variance

The **Variance** of a r.v. X is the following number:

$$Var(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big).$$

In words, Variance shows the following: in the Mean, how far are the values of a r.v. from its Mean, in the squared distance sense?

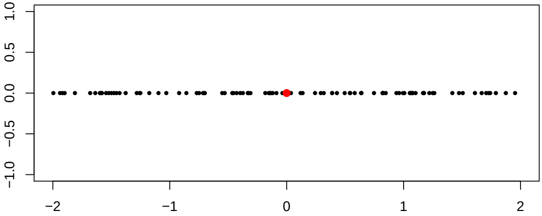
Note: Another important numerical characteristic of the spread around the mean is the Mean Absolute Deviation (MAD):

$$MAD(X) = \mathbb{E}(|X - \mathbb{E}(X)|).$$

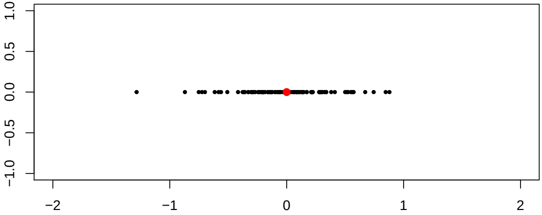
People are mostly using the Variance because of some nice properties we will consider soon.

Variance, Graphically

For comparison: 100 points generated from $X \sim Unif[-2,2]$, $\mathbb{E}(X) = 0$, $Var(X) = \frac{4}{3}$:



100 points generated from $X \sim \mathcal{N}(0, 0.4^2)$, $\mathbb{E}(X) = 0$, $Var(X) = 0.16$:



Variance and SD of a r.v.

The drawback in calculation of the Variance is that it is in squared units. Say, if X is the price of some product, in AMDs, then $\mathbb{E}(X)$ is again in AMD units, but $Var(X)$ will be in AMD^2 units, which is not a convenient measure to report.
So people use the next important characteristic of the Spread, for reporting, the Standard Deviation:

The **Standard Deviation** of a r.v. X is the square root of the Variance:

$$SD(X) = \sigma_X = \sqrt{Var(X)} = \sqrt{\mathbb{E}\Big((X - \mathbb{E}(X))^2\Big)}.$$

Note: Variance is easy to handle with in terms of calculations, and **Standard Deviation** is the measure that is mostly used in reporting.

Variance and SD of a r.v.

When talking about a random quantity, we usually give the Expectation as the typical, average value, as an estimate for that quantity. And with that, we usually report also the Standard Deviation for showing the precision of our estimate.

Example: Say, if we will be informed that the Mean salary at some company is 300K, that gives some information, but is not giving how much deviations we can have from that mean: maybe some persons are getting 10K, others - 1M. But if we will be informed that the Mean salary is 300K with Standard Deviation 40K, then we will be convinced that mostly the salaries are around that 300K. Well, later we will give some quantitative estimates.

Calculation Formula for the Variance

Recall the definition of the Variance:

$$Var(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big).$$

The following formula is widely used in Prob and Stat:

Variance Calculation Formula

$$Var(X) = \mathbb{E}(X^2) - \Big(\mathbb{E}(X)\Big)^2.$$

Proof: OTB

Important Consequence:

$$\mathbb{E}(X^2) = Var(X) + \Big(\mathbb{E}(X)\Big)^2.$$

Calculation of the Variance

Now, let us give explicitly calculation methods for the Variance of a r.v. X:

- If X is Discrete with possible values x_i, and PMF $\mathbb{P}(X = x_i) = p_i$:

$$Var(X) = \sum_i (x_i - \mathbb{E}(X))^2 \cdot p_i = \sum_i x_i^2 p_i - \left(\sum_i x_i p_i\right)^2$$

- If X is Continuous with the PDF $f(x)$:

$$Var(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \left(\int_{-\infty}^{\infty} x f(x)\, dx\right)^2$$

Variance: Example

Example 35.1: Assume

$$X \sim \begin{pmatrix} -1 & 0 & 3 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}$$

Calculate $Var(X)$ and $SD(X)$.

Variance: Example

Example 35.2: Assume

$$X \sim DiscreteUnif(\{x_1, x_2, ..., x_n\}).$$

Find $Var(X)$.

Variance: Example

Example 35.3: Assume

$$X \sim Bernoulli(p).$$

Find $Var(X)$.

Variance: Example

Example 35.4: Assume

$$X \sim Binom(n, p).$$

Find $Var(X)$.

Variance: Example

Example 35.5: Assume

$$X \sim Geom(p).$$

Calculate $Var(X)$.

Variance: Example

Example 35.6: Assume

$$X \sim Pois(\lambda).$$

Calculate $Var(X)$.

Variance: Example

Example 35.7: Assume X is a Continuous r.v. with the following PDF:

$$f(x) = \begin{cases} K \cdot (x+2), & x \in [1,2] \\ 0, & otherwise \end{cases}$$

a. Find K;

b. Calculate $Var(X)$.

Variance: Example

Example 35.8: Assume $X \sim Unif[a, b]$. Find $Var(X)$.

Variance: Example

Example 35.9: Assume $X \sim Exp(\lambda)$.

a. Find $Var(X)$;

b. Use **R** simulations to show the obtained result.

Variance: Example

Example 35.10: Assume $X \sim \mathcal{N}(\mu, \sigma^2)$. Show that $Var(X) = \sigma^2$.

Variance: Example

Example 35.11: Assume I am playing an American Roulette. The board has 37 numbers, from 0 to 36, and an additional slot 00. The slots 0 and 00 are in green, half of the other numbers are in black, and the other half are in red. I have two strategies:

S1. to bet on the color, black/red. If the ball will stop at a pocket with the same color, I will get $1\$$, otherwise I will loose $1\$$;

S2. to do a straight bet: to bet on a number from 0 to 36. If the ball will stop at the pocket with my chosen number, I will get $35\$$, otherwise, I will lose $1\$$.

a. Which strategy is more profitable?

b. Which is more riskier?

Properties of the Variance

Properties of a Variance and a SD

Assume X is a r.v. Then

- $Var(X) \geq 0$ and $SD(X) \geq 0$;
- $Var(X) = 0$ iff $X = Const$ almost surely;
- If X and Y are ID, then $Var(X) = Var(Y)$;
- $Var(X + \alpha) = Var(X)$ for any real number α;
- $SD(X + \alpha) = SD(X)$ for any real number α;
- $Var(\alpha \cdot X) = \alpha^2 \cdot Var(X)$;
- $SD(\alpha \cdot X) = |\alpha| \cdot SD(X)$.

Properties of a Variance and a SD

Note: Assume X is a r.v., and $\alpha, \beta \in \mathbb{R}$. We consider the r.v.

$$Y = \alpha \cdot X + \beta.$$

- The Mean of Y is:

$$\mathbb{E}(Y) = \mathbb{E}(\alpha \cdot X + \beta) = \alpha \cdot \mathbb{E}(X) + \beta;$$

- The Variance of Y is:

$$Var(Y) = Var(\alpha \cdot X + \beta) = \alpha^2 \cdot Var(X).$$

Properties of a Variance and a SD

For a transformed r.v., $Y = g(X)$, in general, there is no nicer formula to calculate the Variance, than

$$Var(Y) = Var(g(X)) = \mathbb{E}\Big(g(X)^2\Big) - \Big[\mathbb{E}(g(X))\Big]^2.$$

Another important property is the calculation of the Variance of a sum of several r.v.s. We will give the general formula later, but one important case is the following:

Variance of a Sum of Independent RVs

Assume X and Y are Independent r.v.s, $X \perp\!\!\!\perp Y$. Then

$$Var(X + Y) = Var(X) + Var(Y).$$

Properties of a Variance and a SD

Note: Important is to remember that

- For **any** r.v.s X, Y (with finite Expectation), both independent, and dependent,

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y);$$

- If X and Y are **Independent**, then

$$Var(X+Y) = Var(X) + Var(Y)$$

  but this is NOT TRUE for any r.v.s (later we will talk about in which cases this property holds);

- And, if X and Y are **Independent**, then

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Variance Properties: Example

Example 35.12: Assume $X, Y \sim Bernoulli(0.5)$ and $X \perp\!\!\!\perp Y$. Calculate

$$Var\big(2X - 3Y - \mathbb{E}(Y) + 5\big).$$

Variance Properties: Example

Example 35.13: Is it true that if $X \perp\!\!\!\perp Y$, then

$$Var(X - Y) = Var(X) - Var(Y),$$

in general?

Variance Properties: Example

Example 35.14: Assume $X \perp\!\!\!\perp Y$. Simplify

$$Var\big(X \cdot Y\big).$$

Variance Properties: Example

Example 35.15: Assume $X \sim Exp(1)$.

a. Calculate $\mathbb{E}(X^2)$;
b. Calculate $\mathbb{E}\Big((X+4)^2\Big)$;
c. Calculate $Var(X^2)$;
d. Calculate $Var(X^2+2X)$.

Variance Properties: Example

Example 35.16: Assume $X, Y, Z \sim Unif[0, 1]$. What can be said about the Distribution of

$$S = X + Y + Z?$$

Solution: Well, we have talked about this last time. In general, if we do not know the relationship between X, Y, Z, we cannot say too much. We can only say that

$$\mathbb{E}(S) = \mathbb{E}(X) + \mathbb{E}(Y) + \mathbb{E}(Z) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}.$$

And, unfortunately, we cannot calculate $Var(S)$. But, if X, Y, Z are independent, then we can calculate the Variance of S easily:

$$Var(S) = Var(X) + Var(Y) + Var(Z) = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{1}{4}.$$

Well, this is easier that to calculate the Distribution of S, then to calculate the Variance.

Variance Properties: Example

Example 35.17: Assume $X_1, X_2, ..., X_n$ are IID r.v.s with

$$Var(X_1) = \sigma^2.$$

Find

a. $Var(X_1 + X_2 + ... + X_n)$;

b. $Var\left(\dfrac{X_1 + X_2 + ... + X_n}{n}\right)$;

c. Interpret the last result.

Variance Properties

The results obtained in the previous example are important, for Probability and Statistics. Let us summarize: Assume we have IID r.v.s $X_1, X_2, ..., X_n$, with

$$\mathbb{E}(X_1) = \mu \qquad \text{and} \qquad Var(X_1) = \sigma^2.$$

Denote

$$S_n = X_1 + X_2 + ... + X_n \qquad \text{and} \qquad \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Then

$$\mathbb{E}(S_n) = n \cdot \mathbb{E}(X_1), \qquad Var(S_n) = n \cdot Var(X_1);$$

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1), \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}.$$

Standardization

Let us recall that we have introduced the Standardization for Normal r.v.s. Now, we can define the Standardization for any r.v.s.

Assume X is a r.v. with finite Expectation $\mathbb{E}(X) = \mu$ and Variance $Var(X) = \sigma^2$. Then the following r.v. is called the Standardization of X:

$$Z = \frac{X - \mathbb{E}(X)}{SD(X)} = \frac{X - \mu}{\sigma}.$$

Unfortunately, if X is not Normal, then Z is not Normal too. But Z has the following nice properties:

$$\mathbb{E}(Z) = 0 \qquad \text{and} \qquad Var(Z) = 1.$$

Standardization

So, summarizing, for any r.v. X, the Standardization yields to

$$Z = \frac{X - \mathbb{E}(X)}{SD(X)} = \frac{X - \mu}{\sigma}.$$

Now,

- If X is Normal r.v., then $Z \sim \mathcal{N}(0, 1)$;
- If X is any r.v., then we have just Z has some Distribution with

$$\mathbb{E}(Z) = 0 \qquad \text{and} \qquad Var(Z) = 1.$$

Variance Properties: Example

Example 35.18: Assume $X \sim Unif[-1, 3]$. Standardize X.

Variance Properties: Example

Example 35.19: Assume $X_1, X_2, ..., X_n$ are IID r.v.s with

$$\mathbb{E}(X_1) = \mu \qquad \text{and} \qquad Var(X_1) = \sigma^2,$$

and

$$S_n = X_1 + X_2 + ... + X_n \qquad \text{and} \qquad \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Standardize S_n and $\overline{X}_n$.

Numerical Characteristics of Important Distributions

Discrete Distributions

- If $X \sim Bernoulli(p)$, then

$$\mathbb{E}(X) = p \qquad Var(X) = p(1-p)$$

- If $X \sim Binom(n, p)$, then

$$\mathbb{E}(X) = n \cdot p, \qquad Var(X) = n \cdot p(1-p)$$

- If $X \sim Geom(p)$, then

$$\mathbb{E}(X) = \frac{1}{p}, \qquad Var(X) = \frac{1-p}{p^2}$$

- If $X \sim Poisson(\lambda)$, then

$$\mathbb{E}(X) = \lambda, \qquad Var(X) = \lambda.$$

Continuous Distributions

- If $X \sim Unif[a, b]$, then

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad Var(X) = \frac{(b-a)^2}{12};$$

- If $X \sim Exp(\lambda)$, then

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad Var(X) = \frac{1}{\lambda^2};$$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{E}(X) = \mu, \qquad Var(X) = \sigma^2.$$

Variance Properties: Example

Example 35.20: Assume X is a Uniform r.v. with

$$\mathbb{E}(X) = 7 \qquad \text{and} \qquad Var(X) = 3.$$

Find $Var(\sqrt{X})$.

Variance Properties: Example

Example 35.21: Assume X, Y are Jointly Distributed r.v.s. Find $\alpha, \beta \in \mathbb{R}$ such that $\alpha \cdot X + \beta$ is the best approximation for Y in the Mean Squared Error sense, i.e.

$$\mathbb{E}\Big((Y - \alpha \cdot X - \beta)^2 \Big)$$

is minimal.

Note: This Problem is important in Statistics and ML. This is the Linear Regression Problem.