Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Probability Distributions: Expectations and Variance

Introduction to Data Science Algorithms

Dirk Grunwald

Why expectation and variance?

- Random variables are complex - contain a lot of information

Why expectation and variance?

- Random variables are complex - contain a lot of information
- If we want to summarize a R.V. by a single value, the *expectation*, or the location of the "central value" would clearly be useful

Why expectation and variance?

- Random variables are complex - contain a lot of information
- If we want to summarize a R.V. by a single value, the *expectation*, or the location of the “central value” would clearly be useful
- If we want to use two values to describe a R.V, the *variance*, or spread from the central value, would be next logical choice.

Why expectation and variance?

- Random variables are complex - contain a lot of information
- If we want to summarize a R.V. by a single value, the *expectation*, or the location of the “central value” would clearly be useful
- If we want to use two values to describe a R.V, the *variance*, or spread from the central value, would be next logical choice.
- You might debate this and prefer *robust* measures, like median and IQR.

Highlights For Expectation

Suppose X is a R.V's, c is a constant and g is an arbitrary function.

1. def. $E[X] = \sum x_i P(X = x_i) = \sum_i x_i p_i$
2. $E[c] = c$
3. $E[cX] = cE[X]$
4. $E[g(X)] = \sum_i g(a_i) P(X = a_i)$

Highlights for Variance

Suppose X is a R.V's, c is a constant and g is an arbitrary function.

1. def. $Var[X] = E\left[(X - E[X])^2\right]$
2. $Var[c] = 0$
3. $Var[cX + s] = c^2 Var[X]$
4. $Var[X] = E[X^2] - (E[X])^2$

Expectation of The Mean..

Expected Mean

- The *expectation* of a discrete R.V. X taking on values $x_1, x_2 \ldots$ is the number

$$E[X] = \sum_k x_k P[X = x_k] = \sum_k x_k p(x_k)$$

- We also call $E[X]$ the *expected value* or *mean* of X
- Similar form for continious R.V.

$$E[X] = \int_k x_k f(x_k)$$

Expected Mean: Example

- Assume a R.V. X is over only the values $\{1, 2, 3\}$ with corresponding probabilities $\{0.1, 0.6, 0.3\}$.

$$E[X] = \sum_i x_i p_i$$

$$= 1 * 0.1 + 2 * 0.6 + 3 * 0.3$$

$$= 0.1 + 1.2 + 0.9$$

$$= 2.2$$

Expectation for Distribution: Bernoulli

Bernouli Distribution with parameter p

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

Expectation for Distribution: Binomial

$$
\begin{aligned}
E[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{k} k \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} k \frac{(n-1)!}{((n-1)(k-1))!k!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{l=0}^{n-1} \binom{n-1}{l} p^{l} (1-p)^{(n-1)-l} \qquad \text{l = k-1} \\
&= np \sum_{l=0}^{m} \binom{m}{l} p^{l} (1-p)^{m-l} \qquad \text{m = n-1} \\
&= np(p+(1-p))^m \qquad \text{Binomial theorem} \\
&= np
\end{aligned}
$$

Expectation for Distribution: Binomial Mk. II

Bernoulli is

$$E[X] = 1 \times p + 0 \times (1 - p) = p \, .$$

Binomial is just n Bernoulli trails.

Earlier we stated that $E[c\, X] = c\, E[X]$. Let's use that.

$$\begin{aligned} E[X_{\text{binomial}}] &= E[n \times X_{\text{bernoulli}}] \\ &= nE[X_{\text{bernoulli}}] \\ &= np \end{aligned}$$

Those rules...

$$E[c] = c$$

Those rules...

$$E[c] = c$$

$$E[cX] = \sum_i c\ x_i p_i = c \sum_i x_i p_i = c\ E[X]$$

Those rules...

$$E[c] = c$$

$$E[cX] = \sum_i c\, x_i p_i = c \sum_i x_i p_i = c\, E[X]$$

This generalizes to a "change of units" (*e.g.* $F \rightarrow C$)

$$E[rX + s] = \sum_i (r\, x_i + s) p_i = r \sum_i x_i p_i + \sum_i s p_i = rE[X] + s$$

Those rules...

$$E[c] = c$$

$$E[cX] = \sum_i c\, x_i p_i = c \sum_i x_i p_i = c\, E[X]$$

This generalizes to a “change of units” (*e.g.* $F \longrightarrow C$)

$$E[rX + s] = \sum_i (r\, x_i + s) p_i = r \sum_i x_i p_i + \sum_i s p_i = rE[X] + s$$

$$E[g(X)] = \sum_i g(x_i) p_i = \sum_i g(x_i) P(X = x_i)$$

Variance..

Variance

Variance is defined as

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

Variance

Variance is defined as

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

Note that $E[X]$ is a constant; could write it *e.g.* μ.

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

Variance

Variance is defined as

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

Note that $E[X]$ is a constant; could write it *e.g.* μ.

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

$Var[X] >= 0$ because it's a sum of squares.

Variance

Variance is defined as

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

Note that $E[X]$ is a constant; could write it *e.g.* μ.

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

$Var[X] >= 0$ because it's a sum of squares.

$SD[X] = \sqrt{Var[X]}$ – same units as $E[X]$.

Expected Variance

The *variance* of a R.V. X is

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

Expected Variance: Example

The *variance* of a R.V. X is $Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$.

Same R.V. X over only the values $\{1,2,3\}$ with corresponding probabilities $\{0.1, 0.6, 0.3\}$, $E[X] = 2.2$

$$Var[X] = 0.1 \times (1 - 2.2)^2 + 0.6 \times (2 - 2.2)^2 + 0.3 \times (3 - 2.2)^2 = 0.36$$

Expected Variance: Example

The *variance* of a R.V. X is $Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$.

Same R.V. X over only the values $\{1,2,3\}$ with corresponding probabilities $\{0.1, 0.6, 0.3\}$, $E[X] = 2.2$

$$Var[X] = 0.1 \times (1 - 2.2)^2 + 0.6 \times (2 - 2.2)^2 + 0.3 \times (3 - 2.2)^2 = 0.36$$

$$Var[X] = (0.1 * 1^2 + 0.6 * 2^2 + 0.6 * 3^2) - 2.2^2 = 5.20 - 4.84 = 0.36$$

Since the variance is a *sum of squares* it's always positive

Examples for Distribution

Bernouli Distribution with parameter p

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

$$\begin{aligned} Var[X] &= E[(X - E[X])^2] \\ &= E[X^2] - E[X]^2 \\ &= p - p^2 \\ &= p(1-p) \end{aligned}$$

This is because

$$\begin{aligned} E[X^2] &= Pr[X=1] * 1^2 + P[X=0] * 0^2 \\ &= p + 0 = p \end{aligned}$$

Change of unit for Variance

$$Var[r\ X + s] = r^2\, Var[X]$$

Change of unit for Variance

$$Var[r\ X+s]=r^2\,Var[X]$$

$$\begin{aligned}
Var[X+s] &= E[(X+s)^2]-E[X+s]^2 \\
&= E[X^2+2sX+s^2]-E[X]^2-2sE[X]-s^2 \\
&= E[X^2]+E[sX]+E[s^2]-E[X]^2-2sE[X]-s^2 \\
&= E[X^2]+sE[X]+s^2-E[X]^2-2sE[X]-s^2 \\
&= E[X^2]-E[X]^2 \\
&= Var[X]
\end{aligned}$$

Change of unit for Variance

$$Var[r\ X+s]=r^2\,Var[X]$$

$$\begin{aligned}
Var[X+s] &= E[(X+s)^2]-E[X+s]^2 \\
&= E[X^2+2sX+s^2]-E[X]^2-2sE[X]-s^2 \\
&= E[X^2]+E[sX]+E[s^2]-E[X]^2-2sE[X]-s^2 \\
&= E[X^2]+sE[X]+s^2-E[X]^2-2sE[X]-s^2 \\
&= E[X^2]-E[X]^2 \\
&= Var[X]
\end{aligned}$$

Generalize..

$$Var[rX+s] = r^2\,Var[X]$$

Undefined means, COV and a warning..

Undefined Means

- Not every probability distribution will have a defined mean, variance or higher order moment.
- For example Let X assume the values $2, 2^2, 2^3, \ldots, 2^k, \ldots$ with pmf $p(x_k) = p(2^k) = 1/2^k$ for $k = 1, 2, 3, \ldots$

Undefined Means

- Not every probability distribution will have a defined mean, variance or higher order moment.
- For example Let X assume the values $2, 2^2, 2^3, \ldots, 2^k, \ldots$ with pmf $p(x_k) = p(2^k) = 1/2^k$ for $k = 1, 2, 3, \ldots$
- It's a valid probability distribution since..

$$
\begin{aligned}
\sum_{k_1}^{\infty} p(x_k) &= \sum_{k=1}^{\infty} 1/2^k \\
&= 1/2 \sum_{k=0}^{\infty} 1/2^k \\
&= 1/2 * \frac{1}{1 - 1/2} = 1
\end{aligned}
$$

Undefined Means - II

- But the mean is undefined..

$$\begin{aligned}\sum_{k_1}^{\infty} x_k p(x_k) &= \sum_{k=1}^{\infty} 2^k \frac{1}{2^k} \\ &= 1 + 1 + \ldots \\ &= \infty\end{aligned}$$

- This is true of the Cauchy distribution and sometimes the Pareto

Squared Coefficient of Variation

- To understand variance relative to $E[X]$, you have to compare two numbers.

Squared Coefficient of Variation

- To understand variance relative to $E[X]$, you have to compare two numbers.
- People are bad at that.

Squared Coefficient of Variation

- To understand variance relative to $E[X]$, you have to compare two numbers.
- People are bad at that.
- C.O.V. is used to measure the degree of irregularity of a positive random variable ($P[X < 0] = 0$).

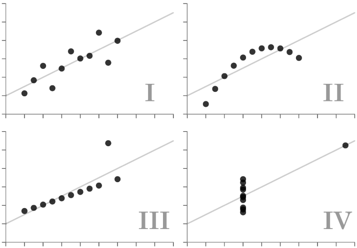
$$C_X^2 = \frac{Var[X]}{E[X]^2}$$

- Or, C.O.V = $\frac{Var[X]}{E[X]^2} = \frac{\sqrt{Var[X]}}{\sqrt{E[X]^2}} = \frac{SD[X]}{|E[X]|}$
- The C.O.V. of an exponential distribution is 1

Don’t Trust Just Numbers

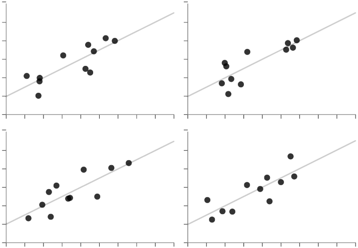
Anscombe’s Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.



Unstructured Quartet

Each dataset here also has the same summary statistics. However, they are not *clearly different* or *visually distinct*.



Don’t Trust Just Numbers

https://www.autodeskresearch.com/publications/samestats