Variance and Standard Deviation

Christopher Croke

University of Pennsylvania

Math 115
UPenn, Fall 2011

Variance

The first first important number describing a probability distribution is the mean or expected value $E(X)$.

Variance

The first first important number describing a probability distribution is the mean or expected value $E(X)$.
The next one is the **variance** $Var(X) = \sigma^2(X)$. The square root of the variance σ is called the **Standard Deviation**.

Variance

The first first important number describing a probability distribution is the mean or expected value $E(X)$.
The next one is the **variance** $Var(X) = \sigma^2(X)$. The square root of the variance σ is called the **Standard Deviation**.
If $f(x_i)$ is the probability distribution function for a random variable with range $\{x_1, x_2, x_3, ...\}$ and mean $\mu = E(X)$ then:

$$Var(X) = \sigma^2 = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + (x_3 - \mu)^2 f(x_3) + ...$$

It is a description of how the distribution "spreads".

Variance

The first first important number describing a probability distribution is the mean or expected value $E(X)$.
The next one is the **variance** $Var(X) = \sigma^2(X)$. The square root of the variance σ is called the **Standard Deviation**.
If $f(x_i)$ is the probability distribution function for a random variable with range $\{x_1, x_2, x_3, ...\}$ and mean $\mu = E(X)$ then:

$$Var(X) = \sigma^2 = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + (x_3 - \mu)^2 f(x_3) + ...$$

It is a description of how the distribution "spreads".
Note $Var(X) = E((X - \mu)^2)$.

Variance

The first first important number describing a probability distribution is the mean or expected value $E(X)$.
The next one is the **variance** $Var(X) = \sigma^2(X)$. The square root of the variance σ is called the **Standard Deviation**.
If $f(x_i)$ is the probability distribution function for a random variable with range $\{x_1, x_2, x_3, ...\}$ and mean $\mu = E(X)$ then:

$$Var(X) = \sigma^2 = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + (x_3 - \mu)^2 f(x_3) + ...$$

It is a description of how the distribution "spreads".
Note $Var(X) = E((X - \mu)^2)$.

The standard deviation has the same units as X. (I.e. if X is measured in feet then so is σ.)

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

The expected value was $\mu = -\frac{2}{15}$.

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

The expected value was $\mu = -\frac{2}{15}$. What is the variance?

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

The expected value was $\mu = -\frac{2}{15}$. What is the variance?

Alternative formula for variance: $\sigma^2 = E(X^2) - \mu^2$.

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

The expected value was $\mu = -\frac{2}{15}$. What is the variance?

Alternative formula for variance: $\sigma^2 = E(X^2) - \mu^2$. Why?

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.
We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

The expected value was $\mu = -\frac{2}{15}$. What is the variance?

Alternative formula for variance: $\sigma^2 = E(X^2) - \mu^2$. Why?

$$E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2)$$

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

The expected value was $\mu = -\frac{2}{15}$. What is the variance?

Alternative formula for variance: $\sigma^2 = E(X^2) - \mu^2$. Why?

$$E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) + E(-2\mu X) + E(\mu^2)$$

Problem: Remember the game where players pick balls from an urn with 4 white and 2 red balls. The first player is paid \$2 if he wins but the second player gets \$3 if she wins. No one gets payed if 4 white balls are chosen.

We have seen that the payout and probabilities for the first player are:

Payout	Probability
2	$\frac{8}{15}$
0	$\frac{1}{15}$
-3	$\frac{6}{15}$

The expected value was $\mu = -\frac{2}{15}$. What is the variance?

Alternative formula for variance: $\sigma^2 = E(X^2) - \mu^2$. Why?

$$E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) + E(-2\mu X) + E(\mu^2)$$

$$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2.$$

This often makes it easier to compute since we can compute μ and $E(X^2)$ at the same time.

This often makes it easier to compute since we can compute μ and $E(X^2)$ at the same time.

Problem: Compute $E(X)$ and $Var(X)$ where X is a random variable with probability given by the chart below:

X	$Pr(X = x)$
1	0.1
2	0.2
3	0.3
4	0.3
5	0.1

This often makes it easier to compute since we can compute μ and $E(X^2)$ at the same time.

Problem: Compute $E(X)$ and $Var(X)$ where X is a random variable with probability given by the chart below:

X	$Pr(X = x)$
1	0.1
2	0.2
3	0.3
4	0.3
5	0.1

We will see later that for X a binomial random variable $\sigma^2 = npq$.

This often makes it easier to compute since we can compute μ and $E(X^2)$ at the same time.

Problem: Compute $E(X)$ and $Var(X)$ where X is a random variable with probability given by the chart below:

X	$Pr(X = x)$
1	0.1
2	0.2
3	0.3
4	0.3
5	0.1

We will see later that for X a binomial random variable $\sigma^2 = npq$.

Problem: What is the variance of the number of hits for our batter that bats .300 and comes to the plate 4 times?

For continuous random variable X with probability density function $f(x)$ defined on $[A, B]$ we saw:

$$E(X) = \int_A^B xf(x)dx.$$

For continuous random variable X with probability density function $f(x)$ defined on $[A, B]$ we saw:

$$E(X) = \int_A^B xf(x)dx.$$

(note that this does not always exist if $B = \infty$.)

For continuous random variable X with probability density function $f(x)$ defined on $[A, B]$ we saw:

$$E(X) = \int_A^B xf(x)dx.$$

(note that this does not always exist if $B = \infty$.)
The variance will be:

$$\sigma^2(X) = Var(X) = E((X - E(X))^2) = \int_A^B (x - E(X))^2 f(x)dx.$$

For continuous random variable X with probability density function $f(x)$ defined on $[A, B]$ we saw:

$$E(X) = \int_A^B x f(x) dx.$$

(note that this does not always exist if $B = \infty$.)
The variance will be:

$$\sigma^2(X) = Var(X) = E((X - E(X))^2) = \int_A^B (x - E(X))^2 f(x) dx.$$

Problem:(uniform probability on an interval) Let X be the random variable you get when you randomly choose a point in $[0, B]$.
a) find the probability density function f.
b) find the cumulative distribution function $F(x)$.
c) find $E(X)$ and $Var(X) = \sigma^2$.

Problem Consider the problem of the age of a cell in a culture. We had the probability density function was $f(x) = 2ke^{-kx}$ on $[0, T]$ where $k = \frac{\ln(2)}{T}$.

a) Find $E(X)$.

Problem Consider the problem of the age of a cell in a culture. We had the probability density function was $f(x) = 2ke^{-kx}$ on $[0, T]$ where $k = \frac{\ln(2)}{T}$.
a) Find $E(X)$.
b) Find σ^2.
It is easier in this case to use the alternative definition of σ^2:

$$\sigma^2 = E(X^2) - E(X)^2 = \int_A^B x^2 f(x)dx - E(X)^2.$$

This holds for the same reason as in the discrete case.

Problem Consider the problem of the age of a cell in a culture. We had the probability density function was $f(x) = 2ke^{-kx}$ on $[0, T]$ where $k = \frac{\ln(2)}{T}$.
a) Find $E(X)$.
b) Find σ^2.
It is easier in this case to use the alternative definition of σ^2:

$$\sigma^2 = E(X^2) - E(X)^2 = \int_A^B x^2 f(x)dx - E(X)^2.$$

This holds for the same reason as in the discrete case.

Problem: Consider our random variable X which is the sum of the coordinates of a point chosen randomly from $[0; 1] \times [0; 1]$? What is $Var(X)$?

Bivariate Distributions

This is the joint probability when you are given two random variables X and Y.

Bivariate Distributions

This is the joint probability when you are given two random variables X and Y.
Consider the case when both are discrete random variables. Then the **joint probability function** $f(x, y)$ is the function:

$$f(x_i, y_j) = Pr(X = x_i, Y = y_j).$$

This is the joint probability when you are given two random variables X and Y.
Consider the case when both are discrete random variables. Then the **joint probability function** $f(x,y)$ is the function:

$$f(x_i, y_j) = Pr(X = x_i, Y = y_j).$$

So of course $f(x_i, y_j) \geq 0$ and the sum over all pairs (x_i, y_j) of $f(x_i, y_j)$ is 1.

This is the joint probability when you are given two random variables X and Y.
Consider the case when both are discrete random variables. Then the **joint probability function** $f(x, y)$ is the function:

$$f(x_i, y_j) = Pr(X = x_i, Y = y_j).$$

So of course $f(x_i, y_j) \geq 0$ and the sum over all pairs (x_i, y_j) of $f(x_i, y_j)$ is 1. Often it is given in the form of a table.

$X \downarrow Y \rightarrow$	1	2	3
0	0.1	0	0.2
2	0.1	0.4	0
4	0.1	0	0.1

This is the joint probability when you are given two random variables X and Y.
Consider the case when both are discrete random variables. Then the **joint probability function** $f(x, y)$ is the function:

$$f(x_i, y_j) = Pr(X = x_i, Y = y_j).$$

So of course $f(x_i, y_j) \geq 0$ and the sum over all pairs (x_i, y_j) of $f(x_i, y_j)$ is 1. Often it is given in the form of a table.

$X \downarrow Y \rightarrow$	1	2	3
0	0.1	0	0.2
2	0.1	0.4	0
4	0.1	0	0.1

What is Pr(X=0,Y=3)?

This is the joint probability when you are given two random variables X and Y.
Consider the case when both are discrete random variables. Then the **joint probability function** $f(x, y)$ is the function:

$$f(x_i, y_j) = Pr(X = x_i, Y = y_j).$$

So of course $f(x_i, y_j) \geq 0$ and the sum over all pairs (x_i, y_j) of $f(x_i, y_j)$ is 1. Often it is given in the form of a table.

$X \downarrow Y \rightarrow$	1	2	3
0	0.1	0	0.2
2	0.1	0.4	0
4	0.1	0	0.1

What is Pr(X=0,Y=3)?

$Pr(X \geq 3, Y \geq 2)$?

This is the joint probability when you are given two random variables X and Y.
Consider the case when both are discrete random variables. Then the **joint probability function** $f(x, y)$ is the function:

$$f(x_i, y_j) = Pr(X = x_i, Y = y_j).$$

So of course $f(x_i, y_j) \geq 0$ and the sum over all pairs (x_i, y_j) of $f(x_i, y_j)$ is 1. Often it is given in the form of a table.

$X \downarrow Y \rightarrow$	1	2	3
0	0.1	0	0.2
2	0.1	0.4	0
4	0.1	0	0.1

What is Pr(X=0,Y=3)?

$Pr(X \geq 3, Y \geq 2)$? $Pr(X = 2)$?

If X and Y are continuous random variables then the **joint probability density function** is a function $f(x, y)$ of two real variables such that for any domain A in the plane:

$$Pr((X, Y) \in A) = \int\int_A f(x, y)dxdy.$$

If X and Y are continuous random variables then the **joint probability density function** is a function $f(x, y)$ of two real variables such that for any domain A in the plane:

$$Pr((X, Y) \in A) = \int \int_A f(x, y) dxdy.$$

We want $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$.

If X and Y are continuous random variables then the **joint probability density function** is a function $f(x, y)$ of two real variables such that for any domain A in the plane:

$$Pr((X, Y) \in A) = \int \int_A f(x, y)dxdy.$$

We want $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$.
Problem: Let f be a joint probability density function (j.p.d.f.) for X and Y where

$$f(x, y) = \begin{cases} c(x + y) & \text{if } x \geq 0, y \geq 0, y \leq 1 - x \\ 0 & \textit{othewise} \end{cases}$$

a) What is c?

If X and Y are continuous random variables then the **joint probability density function** is a function $f(x, y)$ of two real variables such that for any domain A in the plane:

$$Pr((X, Y) \in A) = \int \int_A f(x, y) dxdy.$$

We want $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$.
Problem: Let f be a joint probability density function (j.p.d.f.) for X and Y where

$$f(x, y) = \begin{cases} c(x + y) & \text{if } x \geq 0, y \geq 0, y \leq 1 - x \\ 0 & \textit{othewise} \end{cases}$$

a) What is c?
b) Find $Pr(X \leq \frac{1}{2})$.

If X and Y are continuous random variables then the **joint probability density function** is a function $f(x, y)$ of two real variables such that for any domain A in the plane:

$$Pr((X, Y) \in A) = \int \int_A f(x, y) dxdy.$$

We want $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$.
Problem: Let f be a joint probability density function (j.p.d.f.) for X and Y where

$$f(x, y) = \begin{cases} c(x + y) & \text{if } x \geq 0, y \geq 0, y \leq 1 - x \\ 0 & \textit{othewise} \end{cases}$$

a) What is c?
b) Find $Pr(X \leq \frac{1}{2})$.
c) Set up integral for $Pr(Y \leq X)$.

In some cases X might be discrete and Y continuous (or vice versa).

Bivariate Distributions

In some cases X might be discrete and Y continuous (or vice versa).
In this case we would want $\int_{-\infty}^{\infty} \Sigma_{x_i} f(x_i, y) dy = 1$.

Bivariate Distributions

In some cases X might be discrete and Y continuous (or vice versa).
In this case we would want $\int_{-\infty}^{\infty} \Sigma_{x_i} f(x_i, y) dy = 1$.

The **(cumulative) joint distribution function** for continuous random variables X and Y is

$$F(x, y) = Pr(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) dtds.$$

Bivariate Distributions

In some cases X might be discrete and Y continuous (or vice versa).
In this case we would want $\int_{-\infty}^{\infty} \Sigma_{x_i} f(x_i, y) dy = 1$.

The **(cumulative) joint distribution function** for continuous random variables X and Y is

$$F(x, y) = Pr(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) dt ds.$$

(Use sums if discrete.)

In some cases X might be discrete and Y continuous (or vice versa).
In this case we would want $\int_{-\infty}^{\infty} \Sigma_{x_i} f(x_i, y) dy = 1$.

The **(cumulative) joint distribution function** for continuous random variables X and Y is

$$F(x, y) = Pr(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) dt ds.$$

(Use sums if discrete.)

Thus we see that if F is differentiable

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}$$

Bivariate Distributions

What is the probability in a rectangle? I.e.

$$Pr(a \leq X \leq b, c \leq Y \leq d)?$$

What is the probability in a rectangle? I.e.

$$Pr(a \leq X \leq b, c \leq Y \leq d)?$$

$F(b, d) - F(a, d) - F(b, c) + F(a, c).$

Bivariate Distributions

What is the probability in a rectangle? I.e.

$$Pr(a \leq X \leq b, c \leq Y \leq d)?$$

$F(b,d) - F(a,d) - F(b,c) + F(a,c)$.
Given $F(x,y)$ a c.j.d.f. for X and Y we can find the two c.d.f.'s by:

$$F_1(x) = Pr(X \leq x)$$

Bivariate Distributions

What is the probability in a rectangle? I.e.

$$Pr(a \leq X \leq b, c \leq Y \leq d)?$$

$F(b,d) - F(a,d) - F(b,c) + F(a,c)$.
Given $F(x,y)$ a c.j.d.f. for X and Y we can find the two c.d.f.'s by:

$$F_1(x) = Pr(X \leq x) = \lim_{y \to \infty} Pr(X \leq x, Y \leq y)$$

Bivariate Distributions

What is the probability in a rectangle? I.e.

$$Pr(a \leq X \leq b, c \leq Y \leq d)?$$

$F(b, d) - F(a, d) - F(b, c) + F(a, c)$.
Given $F(x, y)$ a c.j.d.f. for X and Y we can find the two c.d.f.'s by:

$$F_1(x) = Pr(X \leq x) = \lim_{y \to \infty} Pr(X \leq x, Y \leq y) = \lim_{y \to \infty} F(x, y).$$

Bivariate Distributions

What is the probability in a rectangle? I.e.

$$Pr(a \le X \le b, c \le Y \le d)?$$

$F(b,d) - F(a,d) - F(b,c) + F(a,c)$.
Given $F(x,y)$ a c.j.d.f. for X and Y we can find the two c.d.f.'s by:

$$F_1(x) = Pr(X \le x) = \lim_{y\to\infty} Pr(X \le x, Y \le y) = \lim_{y\to\infty} F(x,y).$$

Similarly

$$F_2(y) = \lim_{x\to\infty} F(x,y).$$