Expectation

Statistics for Data Science
CSE357 - Fall 2021

Arithmetic Mean

$$X = \left[2, 3, 3, 4, 4, 4, 5, 5, 6 \right]$$

Arithmetic Mean

$$X = \{2, 3, 3, 4, 4, 4, 5, 5, 6\}$$

$$\frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}$$

An Alternative to Calculate the Mean?

$$X = [\,2, 3, 3, 4, 4, 4, 5, 5, 6\,]$$

$$\frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}$$

An Alternative to Calculate the Mean?

$$X = \left[\, 2,\ \underbrace{3,\ 3},\ \underbrace{4,\ 4,\ 4},\ \underbrace{5,\ 5},\ 6 \,\right] \quad \frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}$$

An Alternative to Calculate the Mean?

$$X = \left[\, 2, \underbrace{3, 3}, \underbrace{4, 4, 4}, \underbrace{5, 5}, 6 \,\right] \qquad \frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}$$

$$2 * \frac{1}{9} + 3 * \frac{2}{9} + 4 * \frac{3}{9} + 5 * \frac{2}{9} + 6 * \frac{1}{9}$$

An Alternative to Calculate the Mean?

$$X = \left[\, 2,\ \underbrace{3,\ 3,}\ \underbrace{4,\ 4,\ 4,}\ \underbrace{5,\ 5,}\ 6 \,\right] \quad \frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}$$

$$2 * \frac{1}{9} + 3 * \frac{2}{9} + 4 * \frac{3}{9} + 5 * \frac{2}{9} + 6 * \frac{1}{9}$$

$$= 2 * \frac{|x = 2|}{|X|} + 3 * \frac{|x = 3|}{|X|} + \ldots$$

An Alternative to Calculate the Mean?

$$X = [\, 2, \underbrace{3, 3}, \underbrace{4, 4, 4}, \underbrace{5, 5}, 6 \,] \qquad \frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}$$

$$2 * \frac{1}{9} + 3 * \frac{2}{9} + 4 * \frac{3}{9} + 5 * \frac{2}{9} + 6 * \frac{1}{9}$$

$$= 2 * \frac{|x = 2|}{|X|} + 3 * \frac{|x = 3|}{|X|} + \ldots$$

$$= \sum_{v \in X} v * \frac{|x = v|}{|X|}$$

Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?

Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?

Formally: The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?

Formally: The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

"expectation" "mean" "first moment"

Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?

Formally: The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x\,dF(x) = \begin{cases} \sum_x x\,f(x) & \text{if } X \text{ is discrete} \\ \int x\,f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

"expectation" "mean" "first moment"

Alternative Conceptualization: If I had to summarize a distribution with only one number, what would do that best?
(the average of a large number of randomly generated numbers from the distribution)

Expectation

the PDF, $f(x)$, is the derivative of the CDF, $F(x)$.

Conceptually: Approximate … en the distr … no other information: what value should I expect?

Formally: The ***expected valu*** … f X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

"expectation" "mean" "first moment"

Alternative Conceptualization: If I had to summarize a distribution with only one number, what would do that best?

(the average of a large number of randomly generated numbers from the distribution)

Expectation

Examples:

$X \sim$ Bernoulli(p):

$X \sim$ Uniform(-3,1):

The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Expectation

Examples:

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} = \binom{1}{x} p^x (1-p)^{(1-x)}$$

***X* ~ Bernoulli(p):**

X ~ Uniform(-3,1):

The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Expectation

Examples:

X ~ Bernoulli(p):

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} = \binom{1}{x} p^x (1-p)^{(1-x)}$$
$$= p^x (1-p)^{(1-x)}$$

X ~ Uniform(-3,1):

The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Expectation

Examples:

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} = \binom{1}{x} p^x (1-p)^{(1-x)}$$
$$= p^x (1-p)^{(1-x)}$$

X ~ Bernoulli(p):

$$\sum_{x \in \{0,1\}} x f(x) = 0(p^0 (1-p)^{(1-0)}) + 1(p^1 (1-p)^{(1-1)})$$

$X \sim$ Uniform(-3,1):

The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Expectation

Examples:

$X \sim$ Bernoulli(p):

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} = \binom{1}{x} p^x (1-p)^{(1-x)}$$
$$= p^x (1-p)^{(1-x)}$$

$$\sum_{x \in \{0,1\}} x f(x) = 0(p^0 (1-p)^{(1-0)}) + 1(p^1 (1-p)^{(1-1)})$$
$$= 1(p^1(1)) = p$$

$X \sim$ Uniform(-3,1):

The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Expectation

Examples:

$X \sim$ Bernoulli(p):

X ~ Uniform(-3,1):

(practice)

$$= \frac{a+b}{2}$$

The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Variance, Second Moment

Conceptually: The expected difference from the mean.

The ***variance*** of X is:

$$V(X) = \sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 dF(x)$$

Variance, Second Moment

$$\int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

Conceptually: The expected difference from the mean.

The ***variance*** of X is:

$$V(X) = \sigma^2 = E(X-\mu)^2 = \int (x-\mu)^2 dF(x)$$

X **~ Bernoulli(p):**

$$p^x(1-p)^{1-x}$$

$$\sum_{x\in\{0,1\}} (x-\mu)^2 (p^x(1-p)^{1-x})$$

Variance, Second Moment

$$\int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

Conceptually: The expected difference from the mean.

The ***variance*** of X is:

$$V(X) = \sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 dF(x)$$

X **~ Bernoulli(p):**

$$p^x(1-p)^{1-x}$$

$$\sum_{x \in \{0,1\}} (x-\mu)^2 (p^x(1-p)^{1-x}) = (0-\mu)^2(p^0(1-p)^{1-0}) + (1-\mu)^2(p^1(1-p)^{1-1})$$

Variance, Second Moment

$$\int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

Conceptually: The expected difference from the mean.

The ***variance*** of X is:

$$V(X) = \sigma^2 = E(X-\mu)^2 = \int (x-\mu)^2 dF(x)$$

X **~ Bernoulli(p):**

$p^x(1-p)^{1-x}$

$$\sum_{x\in\{0,1\}} (x-\mu)^2(p^x(1-p)^{1-x}) = (0-\mu)^2(p^0(1-p)^{1-0}) + (1-\mu)^2(p^1(1-p)^{1-1})$$

$$= \mu^2(1-p) + (1-\mu)^2(p) = p^2(1-p) + (1-p)^2(p)$$

Variance, Second Moment

$$\int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

Conceptually: The expected difference from the mean.

The ***variance*** of X is:

$$V(X) = \sigma^2 = E(X-\mu)^2 = \int (x-\mu)^2 dF(x)$$

***X* ~ Bernoulli(p):**

$p^x(1-p)^{1-x}$

$$\sum_{x\in\{0,1\}} (x-\mu)^2(p^x(1-p)^{1-x}) = (0-\mu)^2(p^0(1-p)^{1-0}) + (1-\mu)^2(p^1(1-p)^{1-1})$$

$$= \mu^2(1-p) + (1-\mu)^2(p) = p^2(1-p) + (1-p)^2(p)$$

$$= p(1-p)(p+1-p) = p(1-p)$$

Distribution	Mean	Variance
Point mass at a	a	0
Bernoulli(p)	p	$p(1-p)$
Binomial(n,p)	np	$np(1-p)$
Geometric(p)	$1/p$	$(1-p)/p^2$
Poisson(λ)	λ	λ
Uniform(a,b)	$(a+b)/2$	$(b-a)^2/12$
Normal(μ,σ^2)	μ	σ^2
Exponential(β)	β	β^2
Gamma(α,β)	$\alpha\beta$	$\alpha\beta^2$
Beta(α,β)	$\alpha/(\alpha+\beta)$	$\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$
t_ν	0 (if $\nu > 1$)	$\nu/(\nu-2)$ (if $\nu > 2$)
χ^2_p	p	$2p$
Multinomial(n,p)	np	see below
Multivariate Normal(μ,Σ)	μ	Σ

discrete

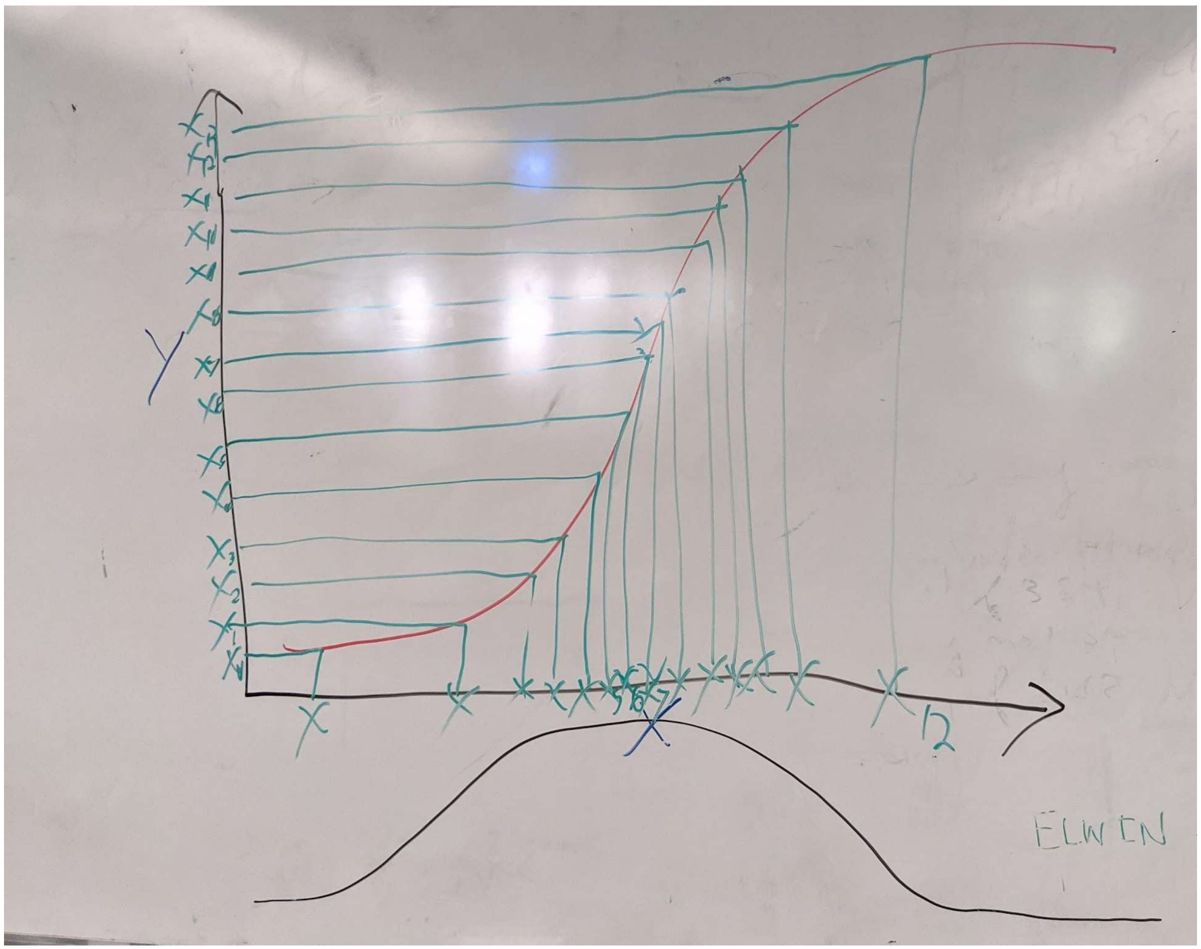
continuous

(Wasserman, 2003)

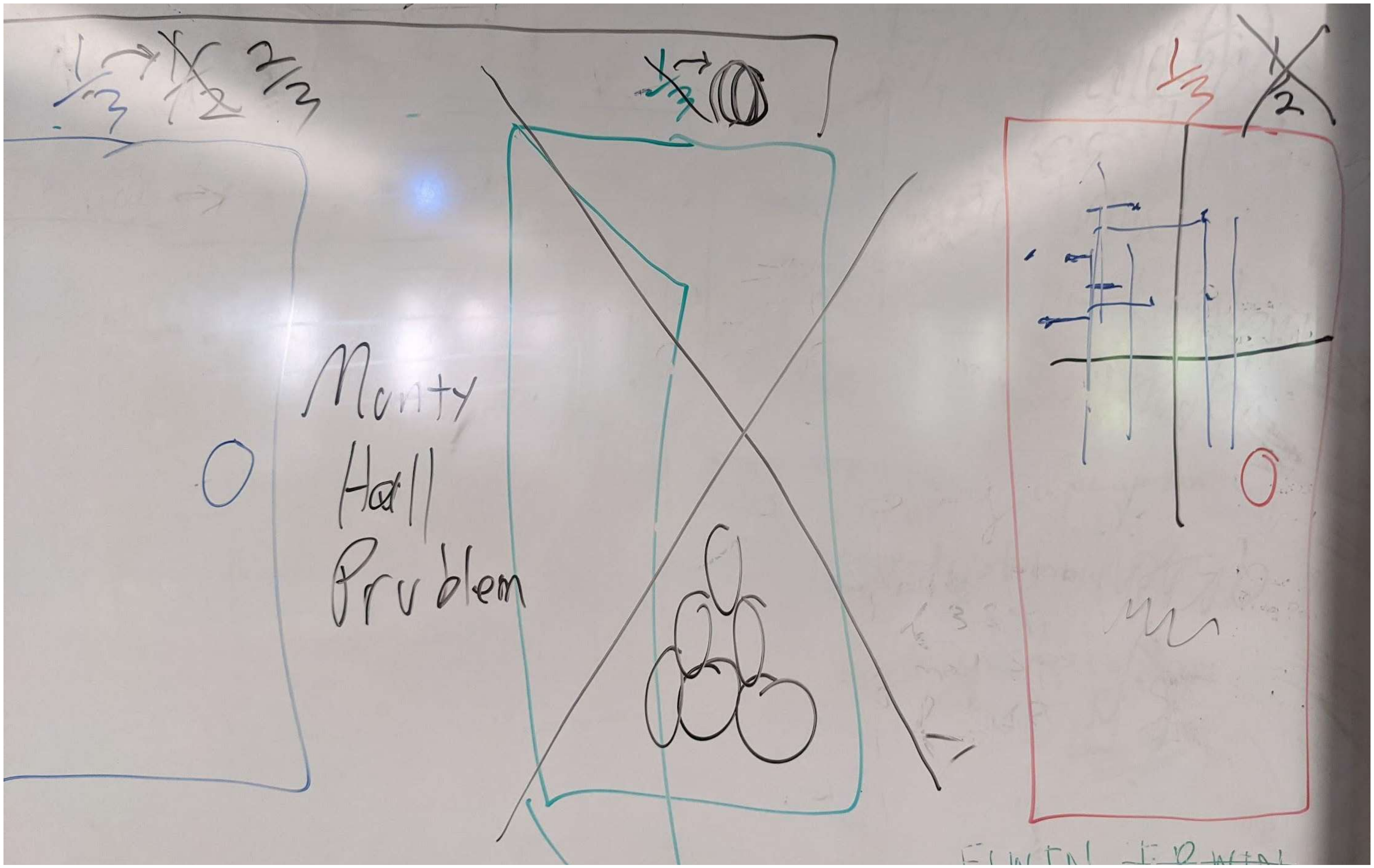
Distribution	Mean	Variance
Point mass at a	a	0
Bernoulli(p)	p	$p(1-p)$
Binomial(n, p)	np	$np(1-p)$
Geometric(p)	$1/p$	$(1-p)/p^2$
Poisson(λ)	λ	λ
Uniform(a, b)	$(a+b)/2$	$(b-a)^2/12$
Normal(μ, σ^2)	μ	σ^2
Exponential(β)	β	β^2
Gamma(α, β)	$\alpha\beta$	$\alpha\beta^2$
Beta(α, β)	$\alpha/(\alpha+\beta)$	$\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$
t_ν	0 (if $\nu > 1$)	$\nu/(\nu-2)$ (if $\nu > 2$)
χ^2_p	p	$2p$
Multinomial(n, p)	np	see below
Multivariate Normal(μ, Σ)	μ	Σ

(Wasserman, 2003)

CDF to PDF trick

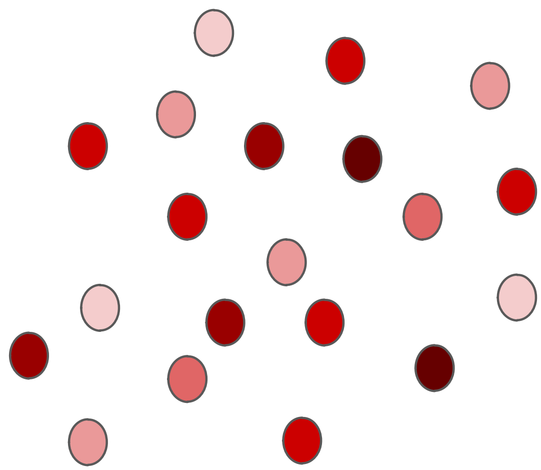


Monty Hall Problem



Population and Samples

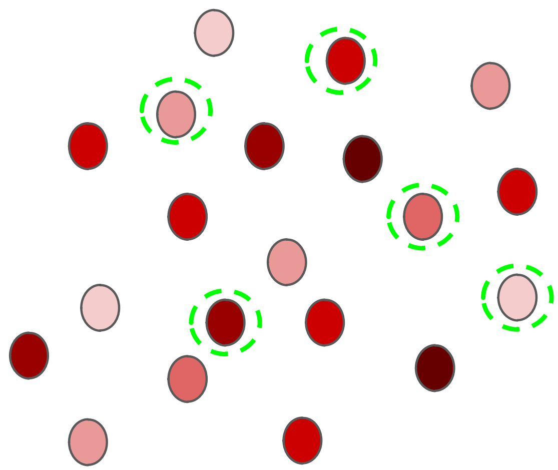
Population



Sample

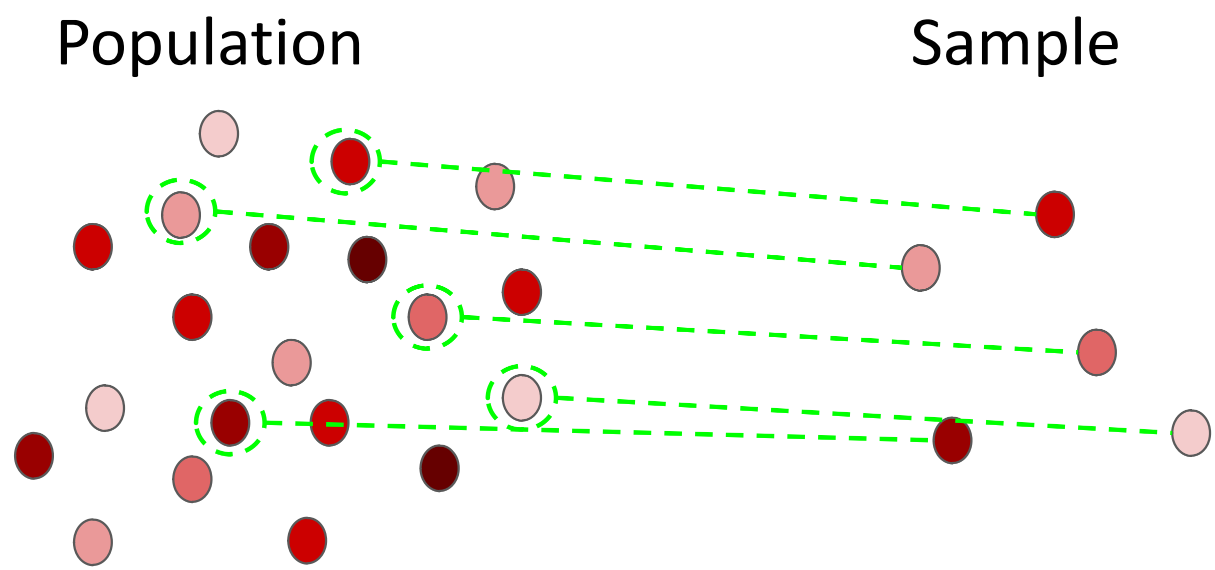
Population and Samples

Population



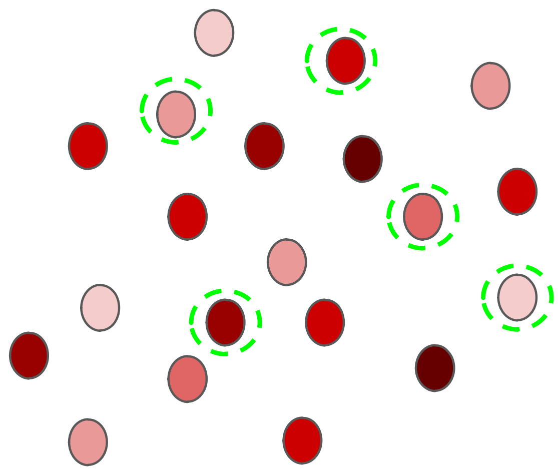
Sample

Population and Samples

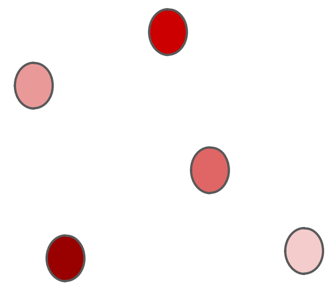


Population and Samples

Population



Sample

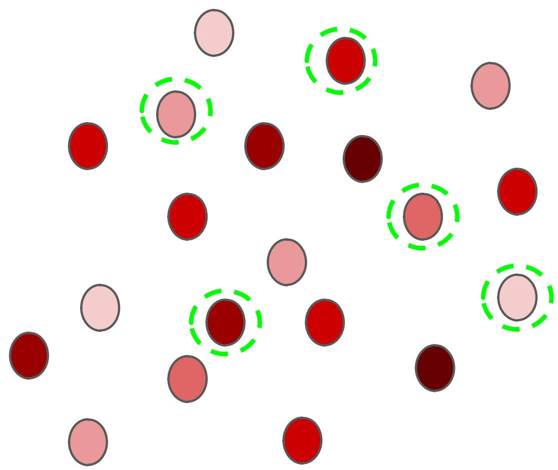


complete: described by pdf of RV
expectation: $E(X)$ or μ
variance: $Var(X)$ or σ^2

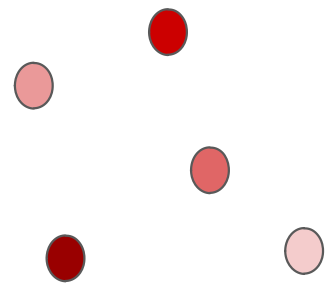
presumed random subset of population
sample mean: $\bar{X}$
sample variance: s^2

Population and Samples

Population



Sample



complete: described by pdf of RV

expectation: $E(X)$ or μ

variance: $Var(X)$ or σ^2

presumed random subset of population

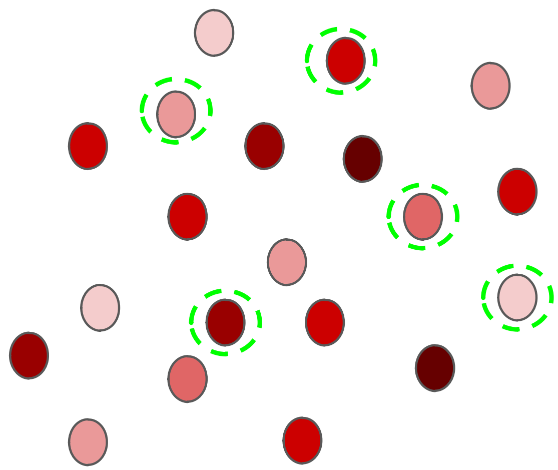
sample mean: $\bar{X}$

sample variance: s^2

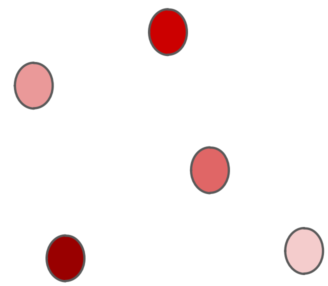
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Population and Samples

Population



Sample



complete: described by pdf of RV

expectation: $E(X)$ or μ

variance: $Var(X)$ or σ^2

presumed random subset of population

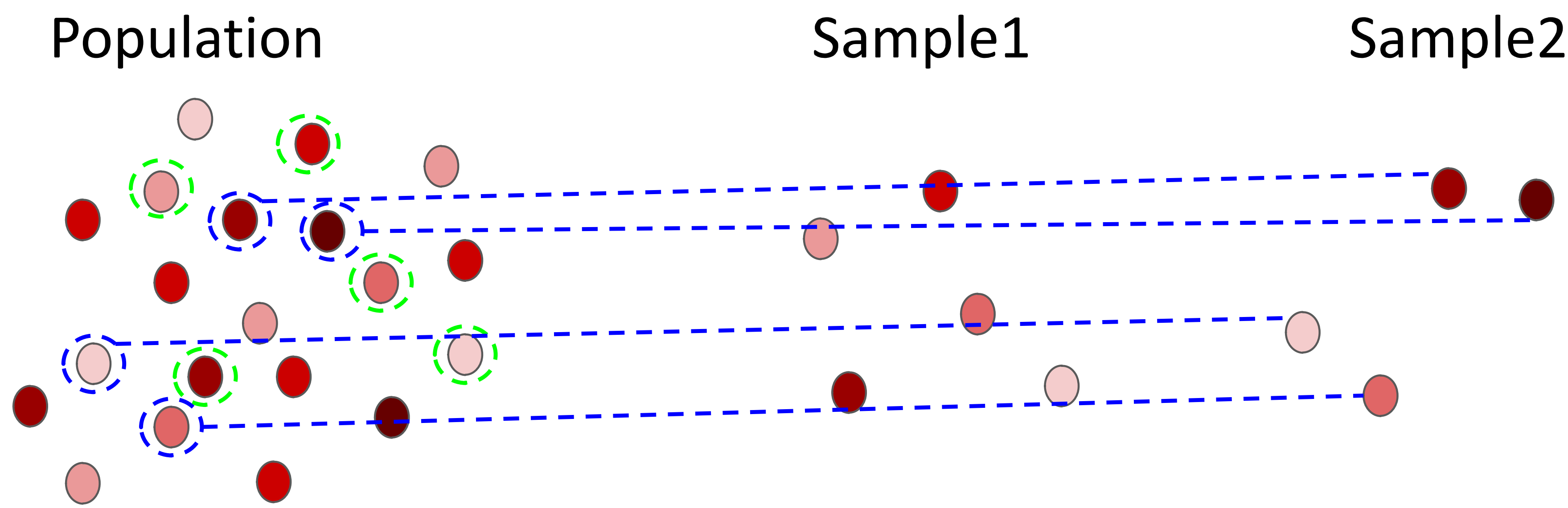
sample mean: $\bar{X}$

sample variance: s^2

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

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

Population and Samples

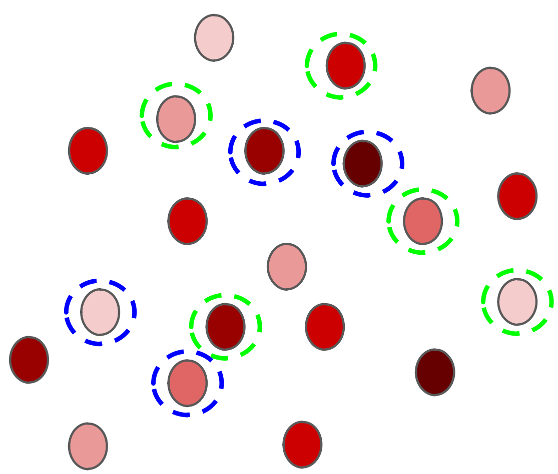


complete: described by pdf of RV
expectation: $E(X)$ or μ
variance: $Var(X)$ or σ^2

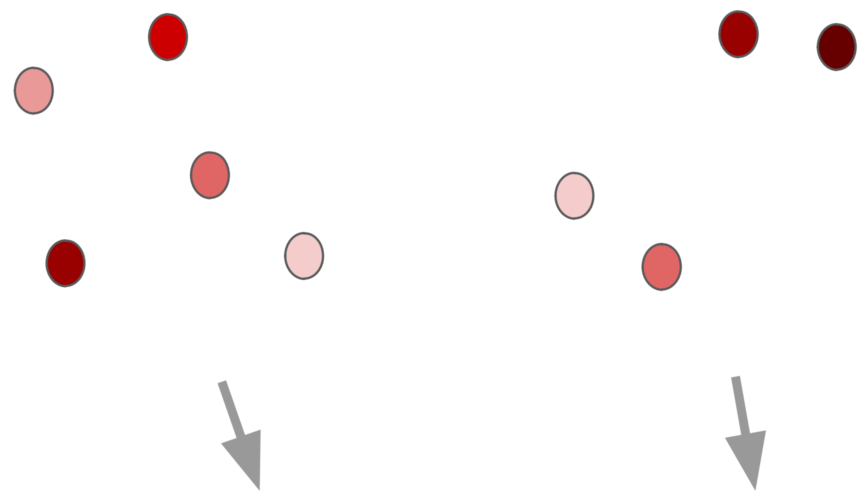
presumed random subset of population
sample mean: $\bar{X}$
sample variance: s^2

Population and Samples

Population



Sample1 **Sample2**



complete: described by pdf of RV

expectation: $E(X)$ or μ

variance: $Var(X)$ or σ^2

presumed random subset of population

	Sample1	Sample2
sample mean:	$\bar{X}_1$	$\bar{X}_2$
sample variance:	s_1^2	s_2^2

Law of Large Numbers

"Weak Law of Large Numbers" (WLLN)

$$if\ X_1, \ldots, X_n are \text{ iid } then \bar{X}_n \rightarrow^p \mu$$

Law of Large Numbers

"Weak Law of Large Numbers" (WLLN)

$$if\ X_1, \ldots, X_n are \text{ iid } then \bar{X}_n \rightarrow^p \mu$$

means $P(|\bar{X} - \mu| > \epsilon) \rightarrow 0$

independent and identically distributed

Law of Large Numbers

"Weak Law of Large Numbers" (WLLN)

$$if\ X_1, \ldots, X_n are \text{ iid } then \underbrace{\bar{X}_n \rightarrow^p \mu}_{}$$

means $P(|\bar{X} - \mu| > \epsilon) \rightarrow 0$

The sample mean converges with the population mean in probability for every $\epsilon > 0$ ($\bar{X}$ is close to μ with high probability)

Central Limit Theorem

Weak LLN: $if\ X_1, \ldots, X_n are\ \text{iid}\ then \bar{X}_n \rightarrow^p \mu$

Can we describe the distribution of $\bar{X} - \mu$?

$\sqrt{n}(\bar{X}_n - \mu)$ converges to a normal distribution.