STATS 100A: Two or More Random Variables

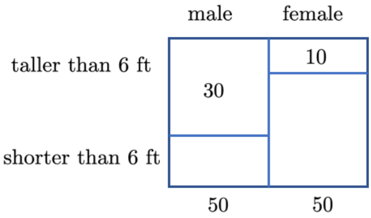
Ying Nian Wu

Department of Statistics
University of California, Los Angeles

Some pictures are taken from the internet.
Credits belong to original authors.

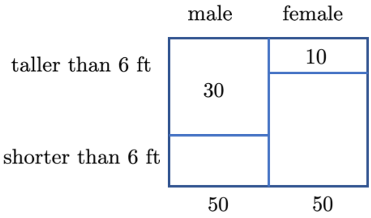
Population

Recall Example 2 in Part 1: Sample a random person from a population of 100 people, 50 males and 50 females. 30 males are taller than 6 ft, 10 females are taller than 6 ft.



Population proportion

Example 2: A male, B tall.



$$P(A) = \frac{|A|}{|\Omega|} = \frac{50}{100} = 50\%.$$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{30+10}{100} = 40\%.$$

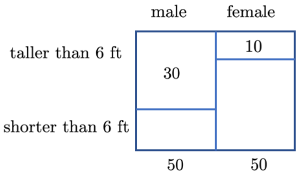
$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{30}{100} = 30\%.$$

Probability = population proportion.

Population proportion

Experiment $\rightarrow$ outcome $\rightarrow$ number

Example 2: A male, B tall.



$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{30}{40} = 75\%.$$

Among tall people, what is the proportion of males?

$$P(B|A) = \frac{|A \cap B|}{|A|} = \frac{30}{50} = 60\%.$$

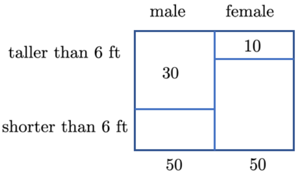
Among males, what is the proportion of tall people?

Conditional probability = proportion within sub-population.

Joint distribution

Example 2: $X \in \{$ male (1), female (0)$\}$, $Y \in \{$ tall (1), short (0)$\}$.

$$p(x, y) = P(X = x, Y = y).$$

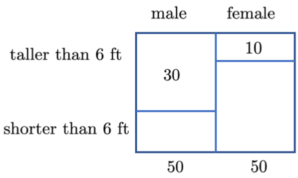


$$p(1,1) = .3, \; p(1,0) = .2,$$
$$p(0,1) = .1, \; p(0,0) = .4.$$

Marginal distribution

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$



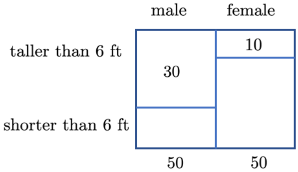
$$P(X = x) = p_X(x) = \sum_y p(x, y).$$

$$P(Y = y) = p_Y(y) = \sum_x p(x, y).$$

Conditional distribution

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$



$$P(X = x|Y = y) = p_{X|Y}(x|y) = p(x, y)/p(y).$$

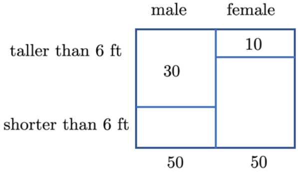
$$P(Y = y|X = x) = p_{Y|X}(y|x) = p(x, y)/p(x).$$

Chain rule: $p(x, y) = p(x)p(y|x) = p(y)p(x|y)$.

Rule of total probability

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x, y) = P(X = x, Y = y).$$

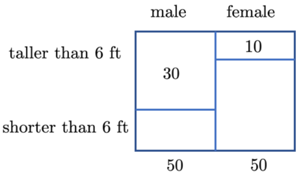


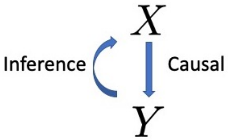
$$p(y) = \sum_x p(x, y) = \sum_x p(x)p(y|x).$$

Bayes rule

Example 2: $X \in \{$ male, female$\}$, $Y \in \{$ tall, short$\}$.

$$p(x,y) = P(X = x, Y = y).$$





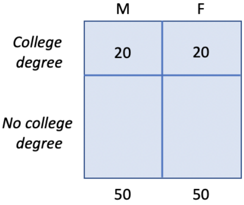
$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_{x'} p(x')p(y|x')}.$$

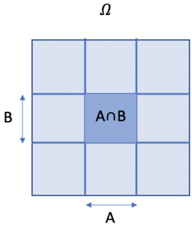
Independence

$$P(A|B) = P(A).$$

$$P(A \cap B) = P(A)P(B).$$

$X \in \{$ male, female$\}$, $Y \in \{$ college, not$\}$





$p(y|x) = p(y).$

$p(x, y) = p(x)p(y|x) = p(x)p(y).$

Recall Example 6: Rare disease example

1% of population has a rare disease.

A random person goes through a test.

If the person has disease, 90% chance test positive.

If the person does not have disease, 90% chance test negative.

If tested positive, what is the chance he or she has disease?

$P(D) = 1\%$.

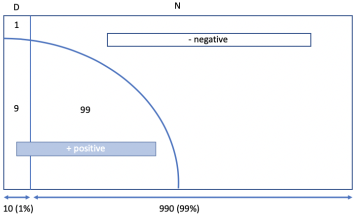
$P(+|D) = 90\%$, $P(-|N) = 90\%$.

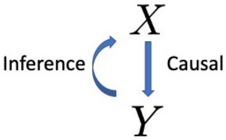
$P(D|+) = ?$

$X \in \{D, N\}$. $Y \in \{+, -\}$.

Reasoning

Example 6: Rare disease example



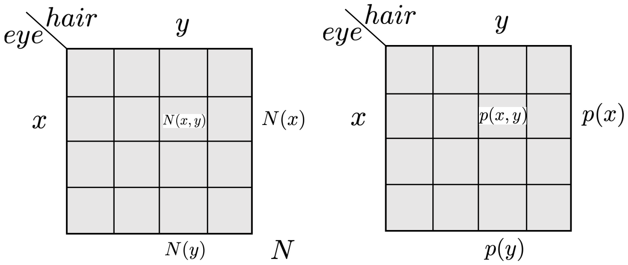


$P(D|+) = \frac{9}{9+99} = \frac{1}{12}$.

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_{x'} p(x')p(y|x')}.$$

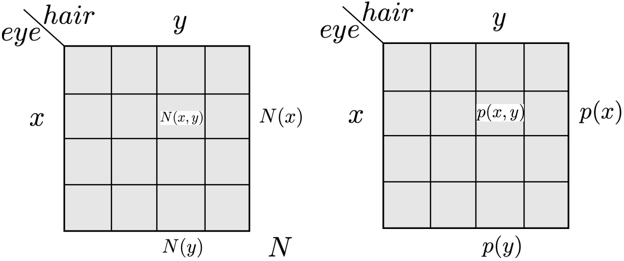
$p(x)$: prior belief. $p(x|y)$: posterior belief.

Discrete distribution



N: number of people in population.
$N(x, y)$: number of people with eye color x and hair color y.
$N(x) = \sum_y N(x, y)$: number of people with eye color x.
$N(y) = \sum_x N(x, y)$: number of people with hair color y.

Joint and marginal

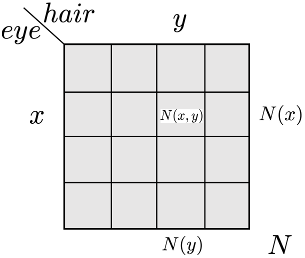


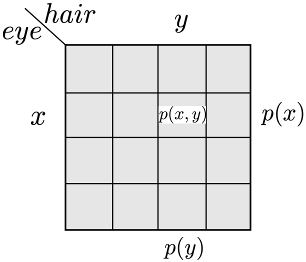
$$p(x,y) = \frac{N(x,y)}{N}.$$

$$p(x) = \frac{N(x)}{N} = \frac{\sum_y N(x,y)}{N} = \sum_y p(x,y).$$

$$p(y) = \frac{N(y)}{N} = \frac{\sum_x N(x,y)}{N} = \sum_x p(x,y).$$

Conditional

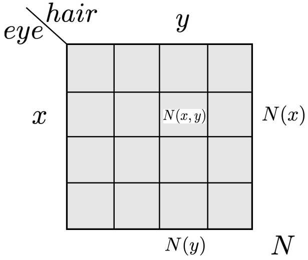


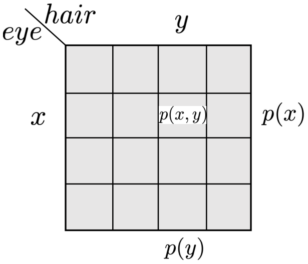


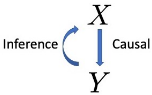
$$p(x|y) = \frac{N(x,y)}{N(y)} = \frac{N(x,y)/N}{N(y)/N} = \frac{p(x,y)}{p(y)}.$$

$$p(y|x) = \frac{N(x,y)}{N(x)} = \frac{N(x,y)/N}{N(x)/N} = \frac{p(x,y)}{p(x)}.$$

Rules

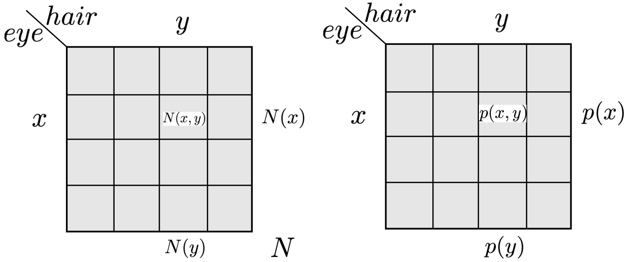






Marginalization: $p(y) = \sum_x p(x,y)$.
Conditioning: $p(x|y) = p(x,y)/p(y)$.
Chain rule: $p(x,y) = p(x)p(y|x)$.

Expectation

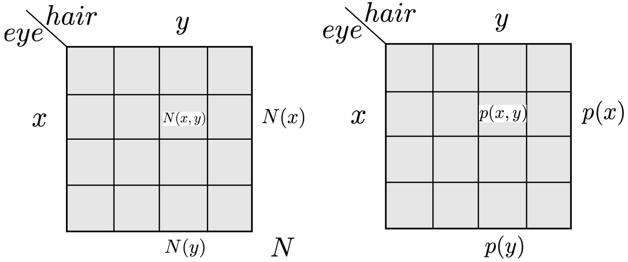


$$\mathbb{E}[h(X,Y)] = \sum_{x,y} h(x,y)p(x,y).$$

Population average or long run average.

$$\frac{1}{N}\sum_{x,y} h(x,y)N(x,y) = \sum_{x,y} h(x,y)\frac{N(x,y)}{N}$$
$$= \sum_{x,y} h(x,y)p(x,y) = \mathbb{E}[h(X,Y)].$$

Expectation



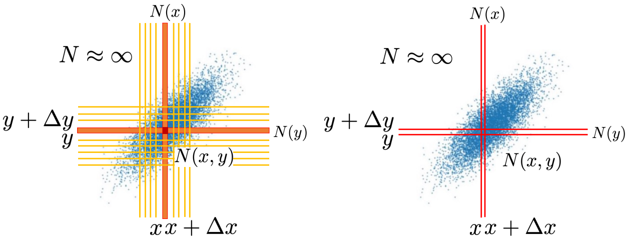
$$\mathbb{E}(X) = \sum_{x,y} xp(x,y) = \sum_x x \sum_y p(x,y) = \sum_x xp(x).$$

same for $\mathbb{E}[h(X)]$.

$$\mathrm{Var}(h(X,Y)) = \mathbb{E}[(h(X,Y) - \mathbb{E}[h(X,Y)])^2].$$

Two continuous random variables

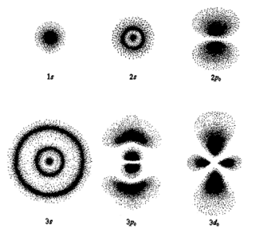
X = height, Y = weight.

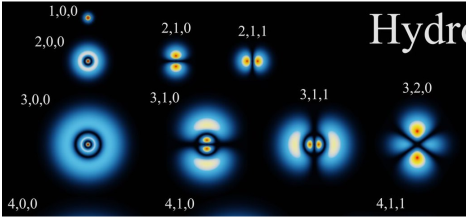


$$f(x, y) = \frac{P(X \in (x, x + \Delta x), Y \in (y, y + \Delta y))}{\Delta x \Delta y} = \frac{N(x, y)/N}{\Delta x \Delta y}.$$

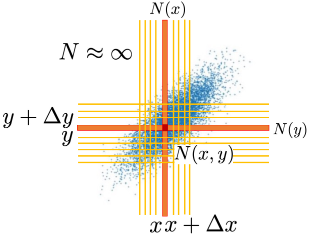
density = probability / size

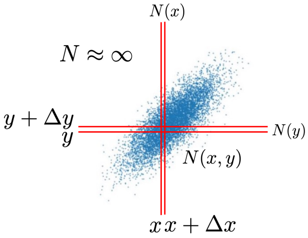
Probability density function





Marginal

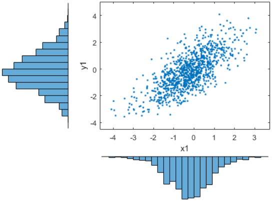


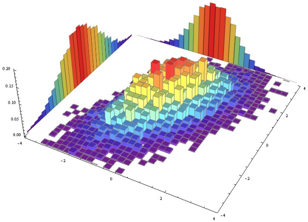


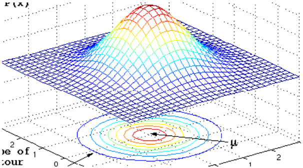
density = prob / size

$$\begin{aligned} f(x) &= \frac{P(X \in (x, x + \Delta x))}{\Delta x} = \frac{N(x)/N}{\Delta x} \\ &= \frac{\sum_y N(x,y)/N}{\Delta x} = \frac{\sum_y f(x,y)\Delta x \Delta y}{\Delta x} = \int f(x,y)dy. \\ f(y) &= \int f(x,y)dx. \end{aligned}$$

Joint and marginal densities

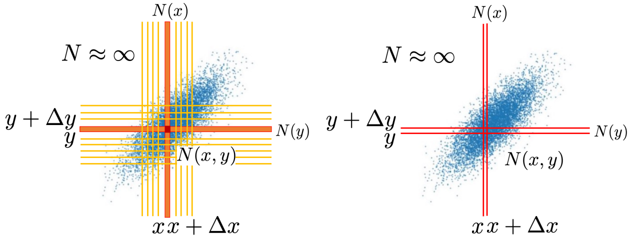






Sample points under the surface, collapse on the plane.

Conditional density

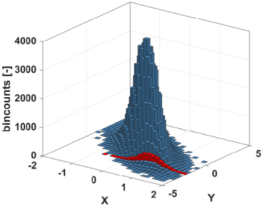


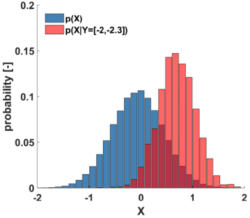
density = prob / size

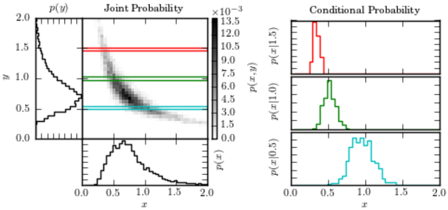
$$f(y|x) = \frac{P(Y \in (y, y+\Delta y) \mid X \in (x, x+\Delta x))}{\Delta y}$$
$$= \frac{N(x,y)/N(x)}{\Delta y} = \frac{N(x,y)/N}{(N(x)/N)\Delta y}$$
$$= \frac{f(x,y)\Delta x \Delta y}{f(x)\Delta x \Delta y} = \frac{f(x,y)}{f(x)}.$$

$f(x|y) = f(x,y)/f(y)$.

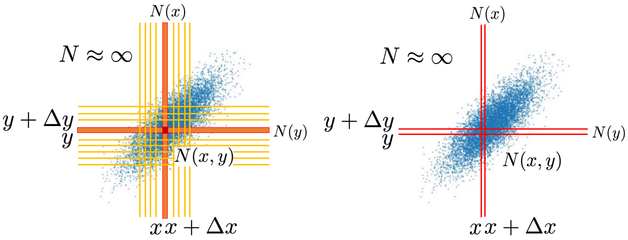
Conditional density

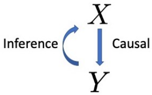






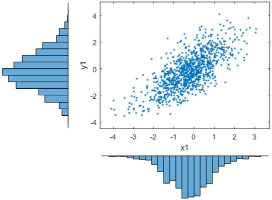
Rules

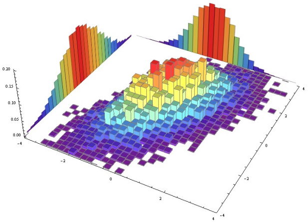




Marginalization: $f(y) = \int f(x, y) dx$.
Normalization (conditioning): $f(x|y) = f(x, y)/f(y)$.
Factorization (chain rule): $f(x, y) = f(x)f(y|x)$.
$f(y|x)$: prediction. $f(x|y)$: inference.

Expectation





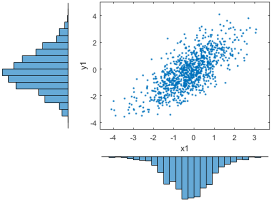
If $(X, Y) \sim p(x, y)$, then

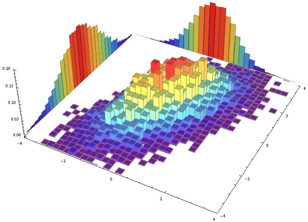
$$\mathbb{E}(h(X, Y)) = \sum_{x} \sum_{y} h(x, y) p(x, y).$$

If $(X, Y) \sim f(x, y)$, then

$$\mathbb{E}(h(X, Y)) = \int \int h(x, y) f(x, y) dx dy.$$

$$\text{Var}(h(X, Y)) = \mathbb{E}[(h(X, Y) - \mathbb{E}[h(X, Y)])^2].$$



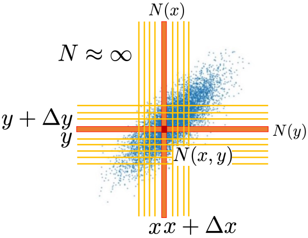


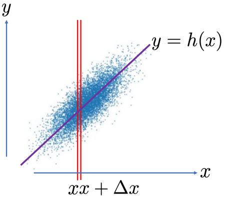
Population average or long run average of $h(X, Y)$.

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i, Y_i) = \frac{1}{n}\sum_{\text{cells}} h(x, y) n f(x, y)\Delta x \Delta y$$

$$\to \int\int h(x, y) f(x, y) dx dy.$$

Conditional expectation and variance





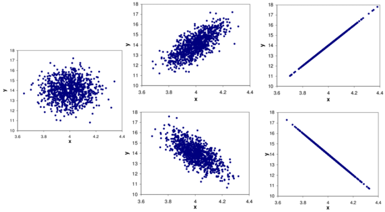
Recall $\mathbb{E}(Y) = \int y f(y) dy$.

$$h(x) = \mathbb{E}[Y|X = x] = \int y f(y|x) dy.$$

Regression, prediction.

$$\text{Var}(Y|X = x) = \mathbb{E}[(Y - h(X))^2 | X = x] = \int (y - h(x))^2 f(y|x) dy.$$

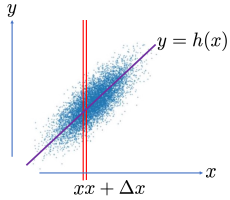
Bivariate Normal



$$X \sim \mathrm{N}(0,1),$$
$$Y = \rho X + \epsilon;\ \epsilon \sim \mathrm{N}(0, 1-\rho^2),\ (|\rho| \leq 1).$$

ϵ is independent of X. Given $X = x$, $Y = \rho x + \epsilon$.

Bivariate Normal



The distribution of points within a vertical slice at x.

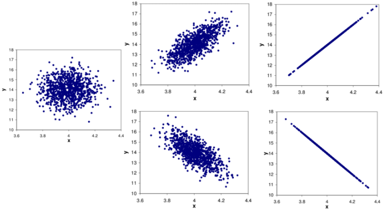
$$\mathbb{E}(Y|X = x) = \mathbb{E}(\rho x + \epsilon) = \rho x.$$

Regression towards the mean ($\rho < 1$), e.g., son's height given father's height.

$$\text{Var}(Y|X = x) = \text{Var}(\rho x + \epsilon) = \text{Var}(\epsilon) = 1 - \rho^2.$$

$$[Y|X = x] \sim \text{N}(\rho x, 1 - \rho^2).$$

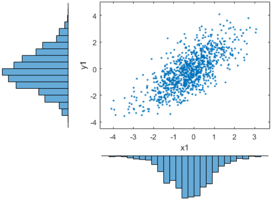
Bivariate Normal

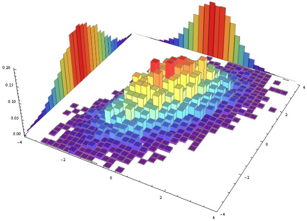


$$
\begin{aligned}
f(x,y) &= f(x)f(y|x) \\
&= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right].
\end{aligned}
$$

symmetric in (x,y)

Covariance



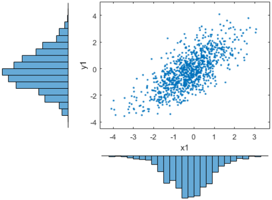


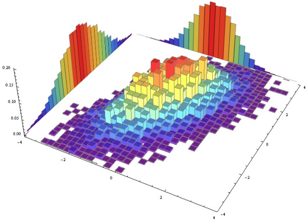
Let $\mu_X = \mathbb{E}(X)$, $\mu_Y = \mathbb{E}(Y)$, we define the covariance

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

It is defined for both discrete and continuous random variables.

Covariance



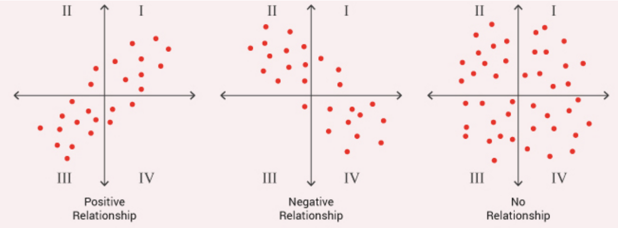


$$(X_i, Y_i) \sim f(x, y),\ i = 1, ..., n.$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i;\ \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

$$\text{Cov}(X, Y) \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

Covariance



$$\text{Cov}(X, Y) \doteq \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}).$$

I, III: $(X_i - \bar{X})(Y_i - \bar{Y}) > 0.$
II, IV: $(X_i - \bar{X})(Y_i - \bar{Y}) < 0.$

Covariance

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \mathbb{E}[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y] \\
&= \mathbb{E}(XY) - \mu_X\mathbb{E}(Y) - \mu_Y\mathbb{E}(X) + \mu_X\mu_Y \\
&= \mathbb{E}(XY) - \mu_X\mu_Y \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}
$$

Clearly, $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$ and $\mathrm{Cov}(Y,Y) = \mathrm{Var}(Y)$.

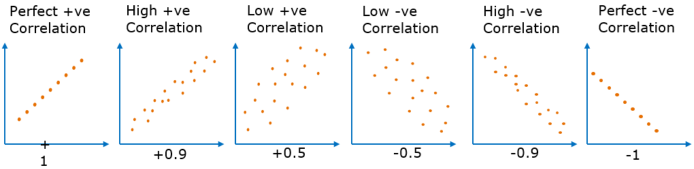
Linearity

$$\begin{aligned}
& \mathrm{Cov}(aX+b, cY+d) \\
= \;& \mathbb{E}[(aX+b-\mathbb{E}(aX+b))(cY+d-\mathbb{E}(cY+d))] \\
= \;& \mathbb{E}[a(X-\mathbb{E}(X))c(Y-\mathbb{E}(Y))] = ac\mathrm{Cov}(X,Y).
\end{aligned}$$

Covariance depends on units (meter/foot, kilogram/pound).

$$\begin{aligned}
& \mathrm{Cov}(X+Y, Z) = \mathbb{E}[(X+Y-\mathbb{E}(X+Y))(Z-\mathbb{E}(Z))] \\
= \;& \mathbb{E}[(X-\mathbb{E}(X)+Y-\mathbb{E}(Y))(Z-\mathbb{E}(Z))] \\
= \;& \mathbb{E}[(X-\mathbb{E}(X))(Z-\mathbb{E}(Z))] + \mathbb{E}[(Y-\mathbb{E}(Y))(Z-\mathbb{E}(Z))] \\
= \;& \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z).
\end{aligned}$$

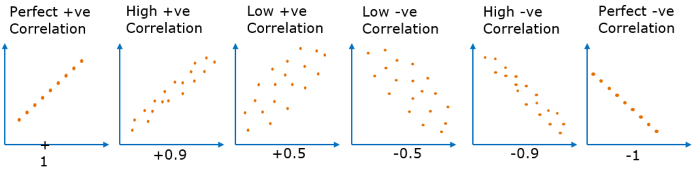
Correlation



Standardize: $X \to (X - \mu_X)/\sigma_X$, $Y \to (Y - \mu_Y)/\sigma_Y$.

$$\text{Cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \text{Corr}(X,Y).$$

Correlation



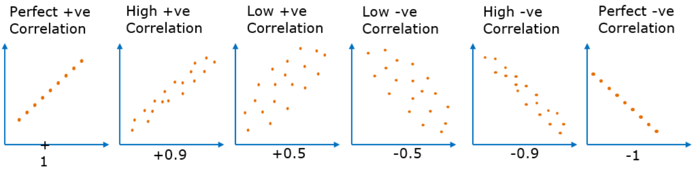
$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}.$$

$$\text{Cov}(X, Y) \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

$$\text{Var}(X) \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2; \ \text{Var}(Y) \doteq \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

$$\text{Corr}(X, Y) \doteq \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

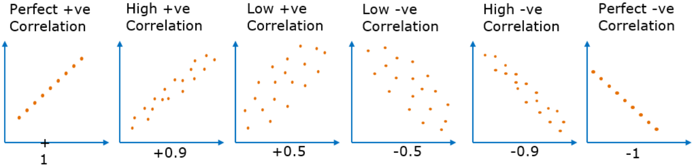
Correlation



Centralize: $\tilde{X}_i = X_i - \bar{X}$; $\tilde{Y}_i = Y_i - \bar{Y}$.

$$\text{Corr}(X, Y) \doteq \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n \tilde{X}_i \tilde{Y}_i}{\sqrt{\sum_{i=1}^n \tilde{X}_i^2}\sqrt{\sum_{i=1}^n \tilde{Y}_i^2}}.$$

Correlation

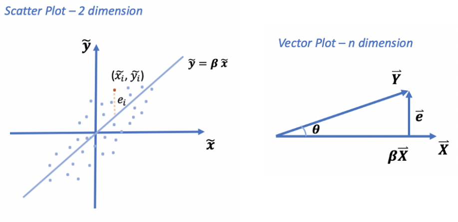


Centralize: $\tilde{X}_i = X_i - \bar{X}$; $\tilde{Y}_i = Y_i - \bar{Y}$.

$$\begin{aligned} \text{Corr}(X, Y) &= \frac{\sum_{i=1}^n \tilde{X}_i \tilde{Y}_i}{\sqrt{\sum_{i=1}^n \tilde{X}_i^2}\sqrt{\sum_{i=1}^n \tilde{Y}_i^2}} \\ &= \frac{\langle \mathbf{X}, \mathbf{Y} \rangle}{|\mathbf{X}||\mathbf{Y}|} = \cos\theta. \end{aligned}$$

Correlation and regression

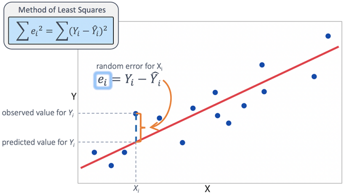
	$\vec{X}$	$\vec{Y}$	$\vec{e}$
1	$\tilde{x}_1$	$\tilde{y}_1$	e_1
⋮	⋮	⋮	⋮
2	$\tilde{x}_i$	$\tilde{y}_i$	e_i
⋮	⋮	⋮	⋮
n	$\tilde{x}_n$	$\tilde{y}_n$	e_n

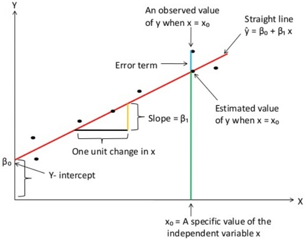


Strength of linear relationship:

$$\frac{\|\mathbf{e}\|^2}{\|\mathbf{Y}\|^2} = \frac{\sum_i e_i^2}{\sum_i (Y_i - \bar{Y})^2} = \sin^2\theta = 1 - \cos^2\theta = 1 - \rho^2.$$

Correlation and regression



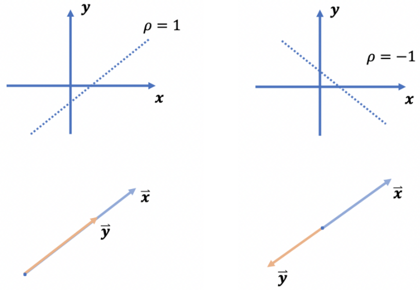


Regression line:

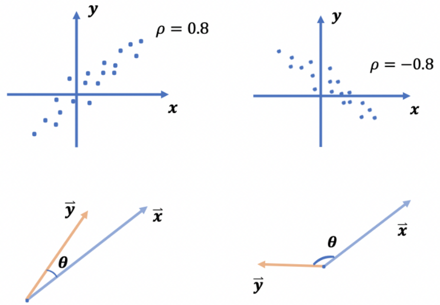
$$\hat{Y} - \bar{Y} = \beta_1 (X - \bar{X}).$$

$$\hat{Y} = \beta_1 X + (\bar{Y} - \beta_1 \bar{X}) = \beta_1 X + \beta_0.$$

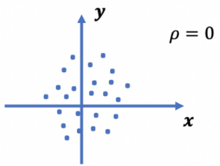
Correlation and regression

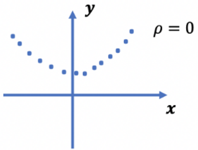


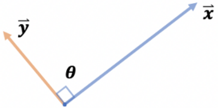
Correlation and regression



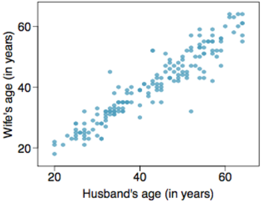
Correlation and regression

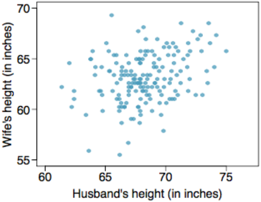


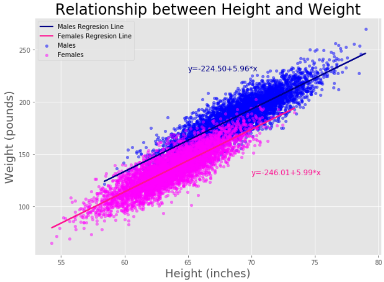




Correlation and regression







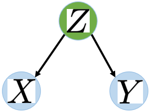
Independence

$P(A \cap B) = P(A)P(B)$.
$p(x,y) = p_X(x)p_Y(y)$; $p(y|x) = p_Y(y)$.
$f(x,y) = f_X(x)f_Y(y)$; $f(y|x) = f_Y(y)$.

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y) \\
&= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p_X(x)p_Y(y) \\
&= \sum_x (x-\mu_X)p_X(x) \sum_y (y-\mu_Y)p_Y(y) \\
&= \left(\sum_x x p_X(x) - \mu_X\right)\left(\sum_y y p_Y(y) - \mu_Y\right) = 0.
\end{aligned}
$$

Conditional independence

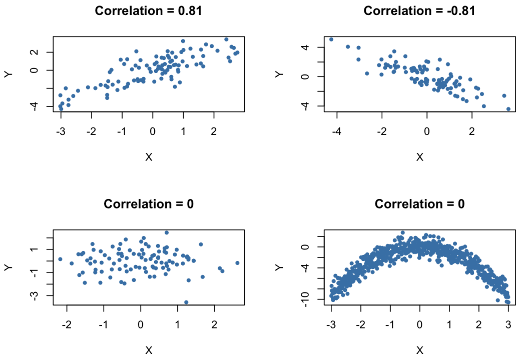
Shared cause: [siblings | parent]



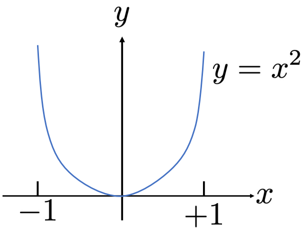
$p(x, y|z) = p(x|z)p(y|z)$
$f(x, y|z) = f(x|z)f(y|z)$
Markov: [future | present, past], [child | parent, grandparent]

$p(y|x, z) = p(y|z)$
$f(y|x, z) = f(y|z)$

Correlation

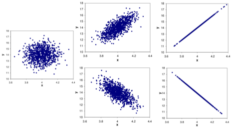


Correlation



Let X be a uniform distribution over $[-1, 1]$. Let $Y = X^2$.
Then X and Y are not independent.
However, $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$, and $\mathbb{E}(X) = 0$. Thus
$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$.

Bivariate normal



$$X \sim \mathrm{N}(0,1),$$
$$Y = \rho X + \epsilon;\ \epsilon \sim \mathrm{N}(0, 1-\rho^2),$$

$$\mathbb{E}(Y) = \mathbb{E}(\rho X + \epsilon) = 0.$$

ϵ and X are independent.

$$\mathrm{Var}(Y) = \mathrm{Var}(\rho X + \epsilon) = \rho^2 \mathrm{Var}(X) + \mathrm{Var}(\epsilon) = 1.$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) = \mathbb{E}[X(\rho X + \epsilon)] = \rho \mathbb{E}(X^2) + \mathbb{E}(X\epsilon) = \rho.$$
$$\mathbb{E}(X\epsilon) = \mathbb{E}(X)\mathbb{E}(\epsilon) = 0.$$

Variance of sum

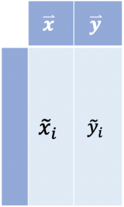
$\mathbb{E}(X+Y) = \sum_x \sum_y (x+y)p(x,y) =$
$\sum_x \sum_y xp(x,y) + \sum_x \sum_y yp(x,y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

$$\begin{aligned} \mathrm{Var}(X+Y) &= \mathbb{E}[((X+Y) - \mu_{X+Y})^2] \\ &= \mathbb{E}[((X-\mu_X) + (Y-\mu_Y))^2] \\ &= \mathbb{E}[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)] \\ &= \mathbb{E}[(X-\mu_X)^2] + \mathbb{E}[(Y-\mu_Y)^2] + 2\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y). \end{aligned}$$

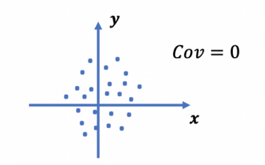
If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$, and

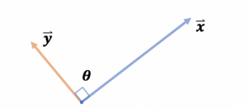
$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Variance of sum

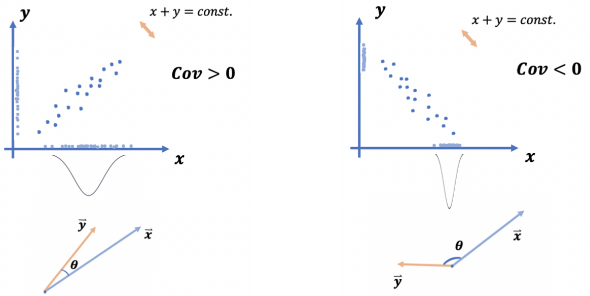


$$\frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i^{\ 2} = Var(X) = \frac{1}{n}|\vec{x}|^2$$





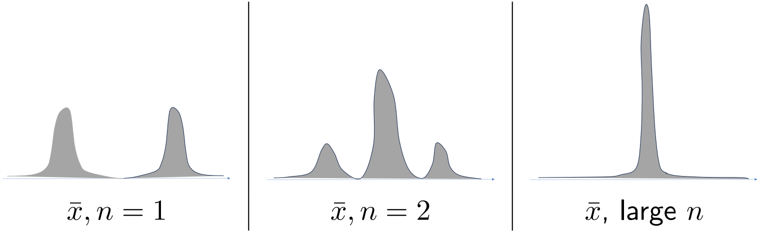
Variance of sum



Average of iid

$X_1, X_2, ..., X_n \sim f(x)$ independently.

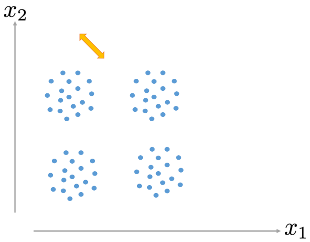
independent and identically distributed, iid



$x_1 \backslash x_2$	small	large
small	small	medium
large	medium	large

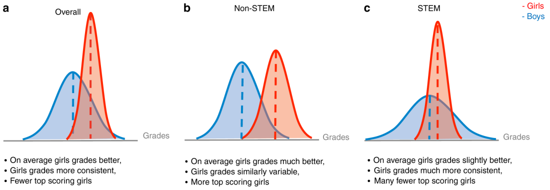
Variance becomes smaller, distribution becomes smoother.

Average of iid



$x_1 \backslash x_2$	small	large
small	small	medium
large	medium	large

Average of iid



Sum and average of iid

$X_i \sim f(x)$, $i = 1, ..., n$, iid: independent and identically distributed.

$$S = \sum_{i=1}^{n} X_i. \; \bar{X} = \frac{S}{n}.$$

$$\mathbb{E}(X_i) = \mu; \; \text{Var}(X_i) = \sigma^2, \; i = 1, ..., n.$$

$$\mathbb{E}(S) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\mu.$$

$$\text{Var}(S) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = n\sigma^2.$$

$$\mathbb{E}(\bar{X}) = \frac{\mathbb{E}(S)}{n} = \mu.$$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(S)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Law of large number

$$\mathbb{E}(\bar{X}) = \frac{\mathbb{E}(S)}{n} = \mu.$$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(S)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0.$$

$$\bar{X} \to \mu, \text{ in probability.}$$

$$P(|\bar{X} - \mu| < \epsilon) \to 1, \; \forall \epsilon > 0.$$

Average $\to$ expectation.

Law of large number

Special case:

$$X = \sum_{i=1}^{n} Z_i,\ Z_i \sim \text{Bernoulli}(p) \text{ iid}.$$

$$\mathbb{E}(X) = np;\ \text{Var}(X) = np(1-p).$$

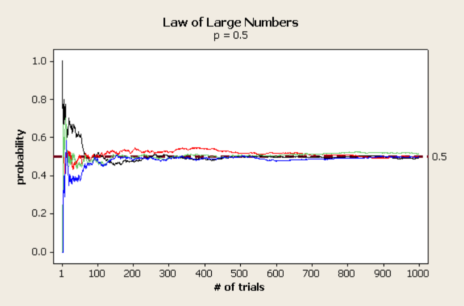
$$\mathbb{E}(X/n) = p;\ \text{Var}(X/n) = p(1-p)/n \to 0.$$

$$X/n \to p, \text{ in probability}.$$

Frequency $\to$ probability.
X/n is average of Z_i. Probability is expectation of Z_i.

Law of large number

Special case:



Keep flipping a fair coin, frequency $\to 1/2$.

Intuition: most of 2^n sequences have frequencies close to 1/2.

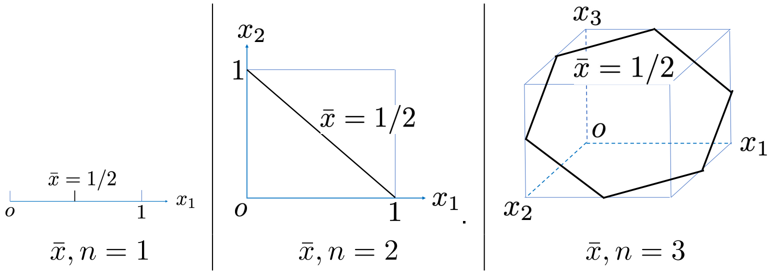
Law of large number

Special case: $X_i \sim \text{Uniform}[0,1]$, iid, $i = 1, ..., n$.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \to \mathbb{E}(X_i) = 1/2.$$

$$P(|\bar{X} - 1/2| < \epsilon) \to 1, \ \forall \epsilon > 0.$$

Intuition: $(X_1, ..., X_i, ..., X_n)$ is a random point in $\Omega = [0,1]^n$, n-dimensional unit cube.
$A = \{(x_1, ..., x_i, ..., x_n) : |\bar{x} - 1/2| < \epsilon\}$ is the central diagonal piece.
$P(A)$ is the volume of A. $P(A) \to 1$.
No matter how small ϵ is, the volume of the central diagonal piece is almost the same as the volume of the whole n-dimensional unit cube Ω.
Most of the points in Ω belong to A. Concentration of measure.


$\bar{x} = 1/2$
o
1
x_1
$\bar{x}, n = 1$
x_2
1
$\bar{x} = 1/2$
o
1
x_1
$\bar{x}, n = 2$
x_3
$\bar{x} = 1/2$
o
x_1
x_2
$\bar{x}, n = 3$

Statistical physics

Most of the points in Ω belong to A. Concentration of measure.

Suppose $(x_1, ..., x_i, ..., x_n)$ describes a physical system, e.g., $n = 10^{23}$ molecules.

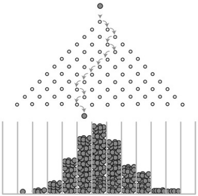
It evolves **deterministically** over time, by traversing with Ω.

Ergodic: it traverses every point in Ω with equal number of visits in the long run.

At any **random moment**, $(x_i, ..., x_i, ..., x_n) \sim \text{Unif}(\Omega)$.

Then most likely it will be in A, with fixed statistical properties (e.g., temperature, pressure, magnetism).

Central limit theorem

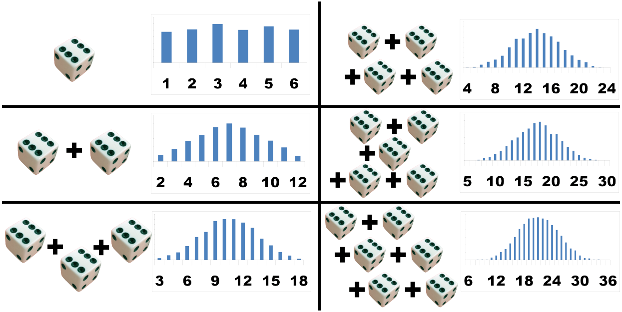


$$X = \sum_{i=1}^{n} \epsilon_i,\ \ \epsilon_i \sim \text{Bernoulli}(1/2) \text{ iid.}$$

$$X \sim \text{Binomial}(n, 1/2).\ \mu = \mathbb{E}(X) = n/2;\ \sigma^2 = \text{Var}(X) = n/4.$$

$$P\left(Z = \frac{X - n/2}{\sqrt{n}/2} = z\right) \doteq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \frac{2}{\sqrt{n}} = f(z)\Delta z.$$

Central limit theorem



Repeat and plot histogram

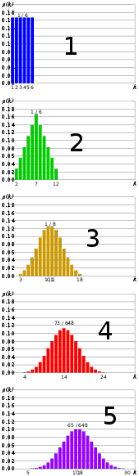
$$S = \sum_{i=1}^{n} X_i.$$

$$\mathbb{E}(X_i) = \mu;\ \mathrm{Var}(X_i) = \sigma^2,\ i = 1, ..., n.$$

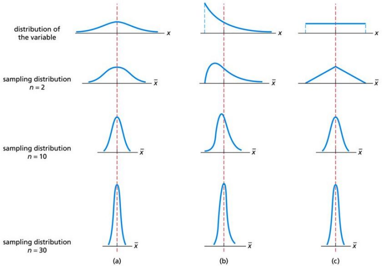
$$S \sim \mathrm{N}(n\mu, n\sigma^2).$$

Central limit theorem

6^n equally likely sequences $\to$ 6^n equally likely sums $\to$ histogram.



Central limit theorem

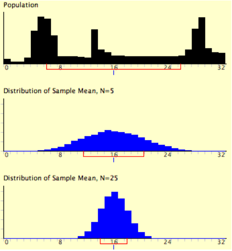


$$S = \sum_{i=1}^{n} X_i. \ \bar{X} = S/n.$$

$$\mathbb{E}(X_i) = \mu; \ \mathrm{Var}(X_i) = \sigma^2, \ i = 1, ..., n.$$

$$S \sim \mathrm{N}(n\mu, n\sigma^2). \ \bar{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$

Central limit theorem



Universal, regardless of the distribution of each X_i.

$$S \sim \mathrm{N}(n\mu, n\sigma^2). \; \bar{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$

$$Z = \frac{S - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1).$$

Probability models

Markov chain: Part 1.
Bayes network, graphical model: Part 1.
Poisson process: Part 2.
Brownian motion: Part 2.

$$X_{t+\Delta t} = X_t + \sigma\sqrt{\Delta t}\epsilon_t,$$

where $\mathbb{E}(\epsilon_t) = 0$, $\mathrm{Var}(\epsilon_t) = 1$, and ϵ_t are iid.
Stochastic differential equation, diffusion

$$X_{t+\Delta t} = X_t + \mu\Delta t + \sigma\sqrt{\Delta t}\epsilon_t,$$

$$dX_t = \mu dt + \sigma dB_t.$$

Imagine 1 million particles moving.

Take home message

As long as you can count (and average)

(1) **Population of equally likely possibilities**

Probability = population proportion

(2) **Large sample of repetitions**

Frequency (sample proportion) $\approx$ probability

(a) **Probability**: population proportion, long run frequency

(b) **Expectation**: population average, long run average

(c) **Conditional**: sub-population, when something happens

Continuous: discretize, infinitesimal analysis