STATS 100A: 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.

Random variables

100A

Ying Nian Wu

Discrete

Continuous

Connection to events:

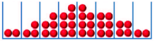
Randomly sample a person ω from a population Ω.

$X(\omega)$: gender of ω, $\Omega \to \{0, 1\}$.
$Y(\omega)$: height of ω, $\Omega \to \mathbb{R}^+$.
$A = \{\omega : X(\omega) = 1\}$. $P(A) = P(X = 1)$. Discrete.
$B = \{\omega : Y(\omega) > 6\}$. $P(B) = P(Y > 6)$. Continuous.
We shall study random variables more systematically.
$\omega \in \Omega$ equally likely, but $X(\omega)$ and $Y(\omega)$ are not necessarily equally likely.

Discrete random variable

Randomly sample a person ω from a population Ω of N people.





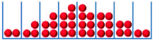
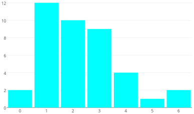
$X(\omega)$: eye color of person ω, $\Omega \to \{1(blue), 2(brown), 3, 4\}$.
$N(x)$ = number of people with eye color x.
Probability mass function, probability distribution, law:

$$X \sim p(x) = P(X = x) = \frac{N(x)}{N}.$$

x	1	2	3	4
$p(x)$	$p(1)$	$p(2)$	$p(3)$	$p(4)$
number	$N(1)$	$N(2)$	$N(3)$	$N(4)$

Discrete random variable

Randomly sample a person ω from a population Ω of N people.

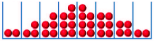
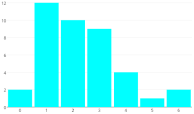




$X(\omega)$: number of siblings of person ω, $\Omega \to \{0, 1, 2, ...\}$.
$N(x)$ = number of people with x siblings.
Probability mass function, probability distribution, law:

$$X \sim p(x) = P(X = x) = \frac{N(x)}{N}.$$

x	0	1	2	3	...
$p(x)$	$p(0)$	$p(1)$	$p(2)$	$p(3)$	...
number	$N(0)$	$N(1)$	$N(2)$	$N(3)$	...

Population average

$N(x)$ = number of people with x siblings.

$$X \sim p(x) = P(X = x) = \frac{N(x)}{N}.$$

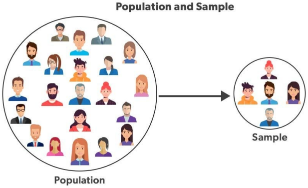
x	0	1	2	3	...
$p(x)$	$p(0)$	$p(1)$	$p(2)$	$p(3)$	...
number	$N(0)$	$N(1)$	$N(2)$	$N(3)$	...

Population average:

$$\mathbb{E}(X) = \frac{1}{N}\sum_{\omega\in\Omega} X(\omega) = \frac{1}{N}\sum_{x} xN(x) = \sum_{x} x\frac{N(x)}{N} = \sum_{x} xp(x).$$

Long run average: N^n reasoning

Randomly sample a person ω from a population Ω of N people. Each person ω carries a number $X(\omega)$.

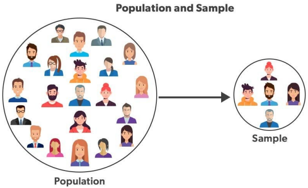


Population average:

$$\mu = \mathbb{E}(X) = \sum_x x p(x).$$

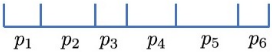
Repeat random sampling n times independently $\rightarrow$
N^n equally likely sequences: Ω_n.
$\bar{X}(\text{sequence})$ = average of the sequence.
$\bar{X} \rightarrow \mu$ in probability as $n \rightarrow \infty$.

Law of large number: N^n reasoning



Repeat random sampling n times independently $\rightarrow$
N^n equally likely sequences: Ω_n.
$\bar{X}(\text{sequence})$ = average of the sequence.
$A = \{\text{sequence} : |\bar{X}(\text{sequence}) - \mu| \leq .01\}$: representative sequences.
$P(A) = \frac{|A|}{N^n} \rightarrow 1$ as $n \rightarrow \infty$.

Die rolling



Randomly throw a point into [0, 1], which bin (1, 2, ..., 6) it falls into?

$\omega \in \Omega = [0, 1]$, population of points (tiny balls), equally likely.

$X(\omega)$ is the bin that ω belongs to, not necessarily equally likely.

Probability mass function, probability distribution, law:

$$X \sim p(x) = P(X = x) = \text{length of bin } x.$$

Population average (each point or tiny ball ω carries a number $X(\omega)$).

$$\mathbb{E}(X) = \sum_x x p(x).$$

Long run average

Independent repetitions:
$p(x)$: how often $X = x$ in the long run (e.g., throw 1 million points into [0, 1]).

x	1	2	3	4	5	6
p(x)	0.1	0.1	0.2	0.2	0.1	0.3

x	1	2	3	4	5	6
#	0.1m	0.1m	0.2m	0.2m	0.1m	0.3m
%	*10%*	*10%*	*20%*	*20%*	*10%*	*30%*

$$average = \frac{(1\times0.1m + 2\times0.1m + 3\times0.2m + 4\times0.2m + 5\times0.1m + 6\times0.3m)}{1m}$$

$$\mathbb{E}(X) = \sum_x xp(x).$$

Expectation of function

Function of random variable

x	1	2	3	4	5	6
p(x)	0.1	0.1	0.2	0.2	0.1	0.3

x	1	2	3	4	5	6
#	0.1m	0.1m	0.2m	0.2m	0.1m	0.3m
%	*10%*	*10%*	*20%*	*20%*	*10%*	*30%*

$$average = \frac{(1\times 0.1m + 2\times 0.1m + 3\times 0.2m + 4\times 0.2m + 5\times 0.1m + 6\times 0.3m)}{1m}$$

x	1	2	3	4	5	6	x
payoff	-$30	-$20	$0	$20	$30	$100	h(x)
	h(1)	*h(2)*	*h(3)*	*h(4)*	*h(5)*	*h(6)*	

longrun average = (-$30)×0.1 + (-$20) ×0.1 + ($0)×0.2 + ($20)×0.2 + ($30)×0.1 + ($100)×0.3

$$\mathbb{E}(h(X)) = \sum_x h(x)p(x).$$

Utility

Utility, reward, value

$$\mathbb{E}(h(X)) = \sum_x h(x)p(x).$$

Offer 1	
x	\$100
p(x)	1

$E(X) = (\$100)\times 1 = \$100$

Offer 2		
x	\$0	\$200
p(x)	1/2	1/2

$E(X) = (\$0)\times\frac{1}{2} + (\$200)\times\frac{1}{2} = \$100$

x: face value	\$0	\$100	\$200
h(x): perceived value	\$0	\$100	\$150

Offer 1: $\mathbb{E}[h(X)] = \$100 \times 1 = \$100.$

Offer 2: $\mathbb{E}[h(X)] = \$0 \times \frac{1}{2} + \$150 \times \frac{1}{2} = \$75.$

Variance

$$\mathbb{E}(X) = \sum_x x p(x) = \mu (= \$0 \times 1/2 + \$200 \times 1/2 = \$100)$$

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x) = \sigma^2 \\ = (\$0 - \$100)^2 \times 1/2 + (\$200 - \$100)^2 \times 1/2 \\ = \$^2 10,000.\end{aligned}$$

Long run average of squared deviation from the mean.

$$SD(X) = \sqrt{\mathrm{Var}(X)} = \sigma (= \$100).$$

Extent of variation from the mean.

Variance

x	1	2	3	4	5	6	x
payoff	-$30	-$20	$0	$20	$30	$100	h(x)
	h(1)	*h(2)*	*h(3)*	*h(4)*	*h(5)*	*h(6)*	

longrun average = (-$30)×0.1 + (-$20) ×0.1 + ($0)×0.2 + ($20)×0.2 + ($30)×0.1 + ($100)×0.3

$$\mathbb{E}(h(X)) = \sum_x h(x)p(x).$$

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

Data

$$\mathbb{E}(X) = \sum_x x p(x) = \mu.$$

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x) = \sigma^2.$$

Long run average of squared deviation from the mean.
Sampling $p(x) \to x_1, ..., x_i, ..., x_n$
(e.g., rolling a die → 2, 1, 6, 5, 3, 2, 5, 4, 3, ...)

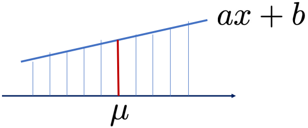
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \to \mathbb{E}(X) = \mu.$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \to \mathrm{Var}(X) = \sigma^2$$

Linear transformation

$$\mathbb{E}(h(X)) = \sum_x h(x)p(x).$$

$$Y = aX + b.$$



$$\mathbb{E}(Y) = \mathbb{E}(aX + b) = \sum_x (ax + b)p(x)$$
$$= \sum_x axp(x) + \sum_x bp(x)$$
$$= a\sum_x xp(x) + b\sum_x p(x) = a\mathbb{E}(X) + b.$$

Data

Sampling $p(x) \to x_1, ..., x_i, ..., x_n$
(e.g., rolling a die $\to$ 2, 1, 6, 5, 3, 2, 5, 4, 3, ...)

$$y_i = ax_i + b.$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \to \mathbb{E}(X) = \mu.$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} (ax_i + b) = a \frac{1}{n} \sum_{i=1}^{n} x_i + b = a\bar{x} + b.$$

Linear transformation

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

$$Y = aX + b.$$

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b.$$

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

$$\begin{aligned}
\mathrm{Var}(aX+b) &= \mathbb{E}[((aX+b) - \mathbb{E}(aX+b))^2] \\
&= \mathbb{E}[(aX + b - (a\mathbb{E}(X) + b))^2] \\
&= \mathbb{E}[(a(X - \mathbb{E}(X)))^2] \\
&= a^2\mathbb{E}[(X - \mathbb{E}(X))^2] = a^2\mathrm{Var}(X).
\end{aligned}$$

Data

Sampling $p(x) \to x_1, ..., x_i, ..., x_n$
(e.g., rolling a die $\to$ 2, 1, 6, 5, 3, 2, 5, 4, 3, ...)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \to \mathbb{E}(X) = \mu.$$

$$y_i = a x_i + b.$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}(a x_i + b) = a\frac{1}{n}\sum_{i=1}^{n} x_i + b = a\bar{x} + b.$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}(a x_i + b - (a\bar{x} + b))^2 = \frac{1}{n}\sum_{i=1}^{n} a^2 (x_i - \bar{x})^2.$$

Short-cut for variance

$$\mu = \mathbb{E}(X).$$

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

$$\begin{aligned}
\mathbb{E}[h(X) + g(X)] &= \sum_x [h(x) + g(x)]p(x) \\
&= \sum_x h(x)p(x) + \sum_x g(x)p(x) \\
&= \mathbb{E}[h(X)] + \mathbb{E}[g(X)].
\end{aligned}$$

Bernoulli

Flip a coin (probability of head is p)
$Z \sim \text{Bernoulli}(p)$
$Z \in \{0, 1\}$, $P(Z = 1) = p$ and $P(Z = 0) = 1 - p$.

$$\mathbb{E}(Z) = 0 \times (1 - p) + 1 \times p = p.$$

$$\begin{aligned} \text{Var}(Z) &= (0 - p)^2 \times (1 - p) + (1 - p)^2 \times p \\ &= p(1 - p)[p + (1 - p)] = p(1 - p). \end{aligned}$$

$$\mathbb{E}(Z^2) = p.$$

$$\text{Var}(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = p - p^2 = p(1 - p).$$

Binomial

Flip a coin (probability of head is p) n times independently.
X = number of heads.
$X \sim \text{Binomial}(n, p)$

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

$\binom{n}{k}$ is the number of sequences with exactly k heads.
$p^k(1-p)^{n-k}$ is the probability of each sequence with k heads.
e.g., $n = 3$,
$P(X = 2) = P(HHT) + P(HTH) + P(THH) = 3p^2(1-p)$.
$p = 1/2$, we have $P(X = k) = \binom{n}{k}/2^n$.

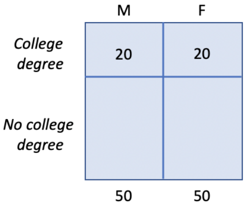
Recall independence

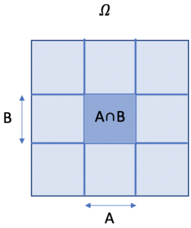
Definition 1:

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

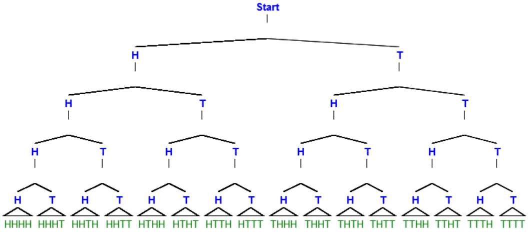
Definition 2:

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





Binomial formula



$$(H+T)^n = \sum_{k=0}^{n} \binom{n}{k} H^k T^{n-k}.$$

$n = 1$, $H + T$.
$n = 2$, $(H+T)(H+T) = HH + HT + TH + TT$.
$n = 3$, above $\times(H+T) =$
$HHH + HHT + HTH + HTT + THH + THT + TTH + TTT$.

Binomial and Bernoulli

$$X = Z_1 + Z_2 + ... + Z_n,$$

where $Z_i \sim \text{Bernoulli}(p)$ independently.

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = np.$$

Due to independence of Z_i, $i = 1, ..., n$,

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(Z_i) = np(1-p).$$

Frequency

X/n is the frequency of heads.

$$\mathbb{E}(X/n) = \mathbb{E}(X)/n = p.$$

$$\text{Var}(X/n) = \text{Var}(X)/n^2 = p(1-p)/n.$$

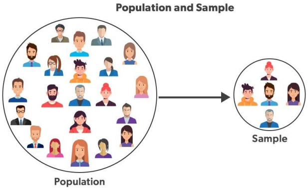
$\text{Var}(X/n) \to 0$ as $n \to \infty$.

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

Law of large number
Probability = long run frequency

Recall survey sampling

Population of N people, M males.
Repeat random sampling n times independently



$\rightarrow N^n$ equally likely sequences.
For a sequence ω, $X(\omega)$ = number of males in ω.
$A_m = \{\omega : X(\omega) = m\}$: sequences with m males.
$|A_m| = \binom{n}{m} M^m (N-M)^{n-m}$. n blanks. Choose m blanks for males, the rest $n-m$ blanks for females. Each male blank has M choices. Each female blank has $N-M$ choices.

Survey sampling

Population of N people. M males.
Sample a person, $p = M/N =$ Prob(male).

$$
\begin{aligned}
P(A_m) = P(X = m) &= \frac{|A_m|}{|\Omega_n|} \\
&= \frac{\binom{n}{m} M^m (N-M)^{n-m}}{N^n} \\
&= \binom{n}{m} p^m (1-p)^{n-m}.
\end{aligned}
$$

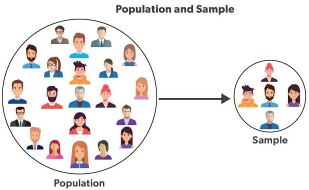
Most sequences are representative, $X/n \approx M/N = p$.

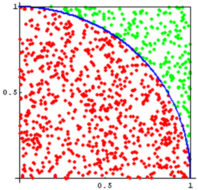
Binomial distribution

Flip a coin n times independently, $p =$ probability of head.

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$x = 0, 1, ..., n.$$

$p(x)$: probability mass function, probability distribution.

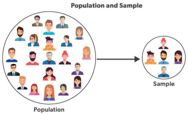




Survey sampling, poll before election, $p = M/N$.
Monte Carlo, $p = \pi/4$.

Law of large number: N^n reasoning

Survey sampling, poll before election, $p = M/N$.



Among all N^n sequences in the hyper-population of sequences Ω_n, let

$$A = \left\{\omega : \left|\frac{X(\omega)}{n} - p\right| \leq .01\right\}.$$

consist of representative sequences.

$$P(A) = \frac{|A|}{|\Omega_n|} = \sum_{x \in [n(p-.01), n(p+.01)]} p(x) \to 1,$$

$X/n \to p$ in probability.
$\mathbb{E}(X/n) = p$: average of $X(\omega)/n$ in Ω_n.
$\mathrm{Var}(X/n) = p(1-p)/n \to 0$: variance of $X(\omega)/n$ in Ω_n.

Binomial expectation

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{n} x p(x) = \sum_{k=0}^{n} k P(X = k) \\
&= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} np \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= \sum_{k'=0}^{n'} np \binom{n'}{k'} p^{k'} (1-p)^{n'-k'} = np.
\end{aligned}
$$

$$k' = k - 1;\ n' = n - 1.$$

Binomial variance

$$\mathbb{E}(X(X-1)) = \sum_{x=0}^{n} x(x-1)p(x) = \sum_{k=0}^{n} k(k-1)P(X=k)$$
$$= \sum_{k=0}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$
$$= \sum_{k=2}^{n} n(n-1)p^2\frac{(n-2)!}{(k-2)!(n-k)!}p^{k-2}(1-p)^{n-k}$$
$$= \sum_{k'=0}^{n'} n(n-1)p^2\binom{n'}{k'}p^{k'}(1-p)^{n'-k'}$$
$$= n(n-1)p^2.$$

$$k' = k-2;\ n' = n-2.$$

Binomial variance

$$\mathbb{E}(X) = np.$$

$$\mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}(X) = n(n-1)p^2.$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= np - np^2 = np(1-p).
\end{aligned}$$

Geometric

$T \sim \text{Geometric}(p)$
T is the number of flips to get the first head, if we flip a coin independently and the probability of getting a head in each flip is p.

$$P(T = k) = (1-p)^{k-1}p.$$

e.g., $T = 1$, H
$T = 2$, TH.
$T = 3$, TTH.
$T = 4$, $TTTH$.
Waiting time.

Geometric expectation

$T \sim \text{Geometric}(p)$

$$
\begin{aligned}
\mathbb{E}(T) &= \sum_{k=1}^{\infty} kP(T=k) \\
&= \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=1}^{\infty} \frac{d}{dq} q^k \\
&= p\frac{d}{dq}\sum_{k=1}^{\infty} q^k = p\frac{d}{dq}\left(\frac{1}{1-q} - 1\right) \\
&= p\frac{1}{(1-q)^2} = \frac{1}{p}.
\end{aligned}
$$

Geometric series

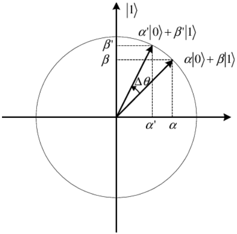
$$
\begin{aligned}
(1-a)(1+a+...+a^m) &= 1+a+...+a^m \\
&\quad -(a+a^2+...+a^m+a^{m+1}) \\
&= 1-a^{m+1}.
\end{aligned}
$$

$$1+a+...+a^m = \frac{1-a^{m+1}}{1-a}.$$

If $|a| < 1$,

$$a^{m+1} \to 0, \ as \ m \to \infty.$$

Quantum bit

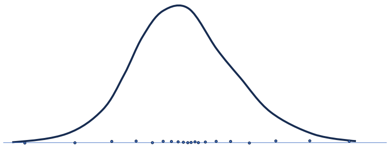


state vector $= \alpha|0\rangle + \beta|1\rangle$.
state vector rotates over time.
squared length $= |\alpha|^2 + |\beta|^2 = 1$ under rotation.
observer: $p(0) = |\alpha|^2$, $p(1) = |\beta|^2$.

Schrodinger cat: P(alive) $= (1/\sqrt{2})^2 = 1/2$.

Continuous random variable: density

Randomly sample a person ω from a population Ω of N people.
$X(\omega)$: height of person ω.
Density or distribution of the N points.

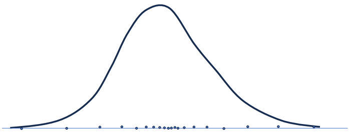


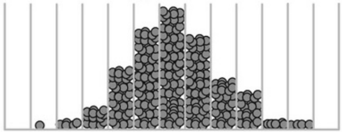
$N(x) =$ number of people in $(x, x + \Delta x)$ (6 ft, 6 ft 1 inch), precision $=$ 1 inch.
Probability density function, probability distribution, law:

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

Mathematical idealization: $N \approx \infty$.

Population scatterplot and histogram





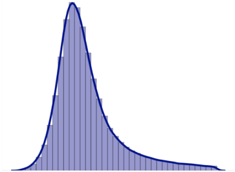
Discretize x-axis into equally spaced bins $(x, x + \Delta x)$, e.g., (6 ft, 6 ft 1 inch), precision = 1 inch.

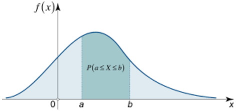
$$P(X \in (x, x + \Delta x)) = \frac{N(x)}{N} = f(x)\Delta x.$$

$f(x)$: height of bin $(x, x + \Delta x)$. $f(x)\Delta x$: area.

$$\sum_x \frac{N(x)}{N} = \sum_x f(x)\Delta x \to \int f(x)dx = 1.$$

Region under curve



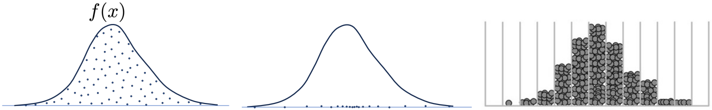


Randomly throw a point ω into the region Ω below curve $f(x)$.
Ω: population of points (tiny squares or balls).
Let $X = X(\omega)$ be the horizontal coordinate of point ω.

$$P(X \in (x, x + \Delta x)) = f(x)\Delta x.$$

$$P(X \in (a, b)) = \sum_{x \in (a,b)} f(x)\Delta x \to \int_a^b f(x)dx.$$

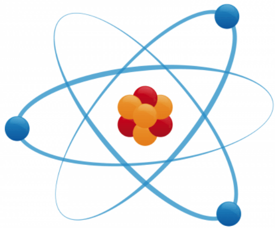
Independent repetitions, sample scatterplot and histogram

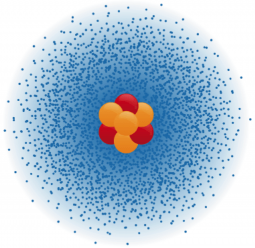


Repeat n times, collapse to x-axis, histogram.
N^n reasoning:
sample scatterplot (random) $\approx$ population scatterplot (fixed),
sample histogram (random) $\approx$ population histogram (fixed).

Point cloud

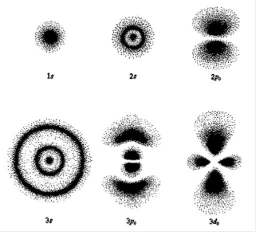
Electron orbits around nucleus: wrong conception
Electron cloud, probability density function, $f(x)$
Wave function $\psi(x)$, evolves over time.
Observer: $f(x) = |\psi(x)|^2$.

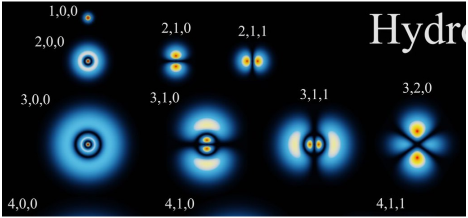




Population

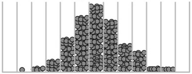
Electron cloud, heat map, prob density

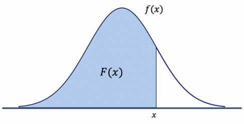


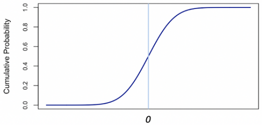


Population of N equally likely possibilities.
Mathematical idealization: $N \approx \infty$.
Prob density = prob mass in the cell / volume of cell.
Observer: $f(x) = |\psi(x)|^2$.

Cumulative density function



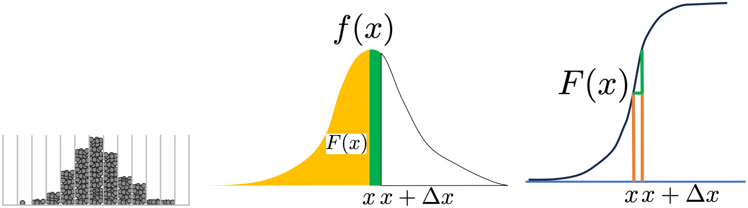




$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx.$$

SAT score $x \rightarrow$ percentile $F(x)$.
Percentage of people below x.

Area and slope



Area:

$$F(x+\Delta x) - F(x) = f(x)\Delta x.$$

Slope:

$$F'(x) = \frac{F(x+\Delta x) - F(x)}{\Delta x} = f(x).$$

Notation:

$$F'(x) = \frac{dF(x)}{dx} = \frac{d}{dx}F(x) = f(x).$$

$$dF(x) = F'(x)dx = f(x)dx$$

Expectation

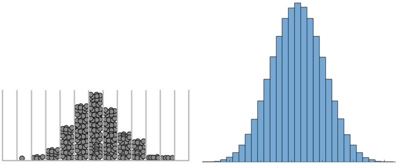
Recall discrete

$$P(X = x) = p(x).$$

$$\mathbb{E}(X) = \sum xP(X = x) = \sum xp(x).$$

Continuous

$$P(X \in (x, x + \Delta x)) = f(x)\Delta x.$$

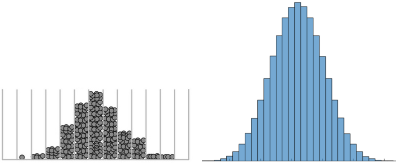


$$\mathbb{E}(X) = \sum xP(X \in (x, x+\Delta x)) = \sum xf(x)\Delta x \to \int xf(x)dx.$$

Population average, long run average, center.

Population average

Population Ω of N people. $X(\omega)$.
$N(x)$: number of people in $(x, x + \Delta x)$.



$$\mathbb{E}(X) = \frac{1}{N} \sum_{\omega} X(\omega) = \frac{1}{N} \sum_{x} xN(x)$$

$$= \sum_{x} x \frac{N(x)}{N} = \sum xP(X \in (x, x + \Delta x))$$

$$= \sum xf(x)\Delta x \to \int xf(x)dx.$$

Expectation of function

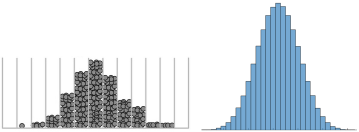
Recall discrete

$$P(X = x) = p(x).$$

$$\mathbb{E}[h(X)] = \sum h(x)P(X = x) = \sum h(x)p(x).$$

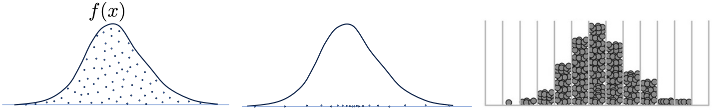
Continuous

$$P(X \in (x, x + \Delta x)) = f(x)\Delta x.$$



$$\begin{aligned}\mathbb{E}[h(X)] &= \sum h(x)P(X \in (x, x + \Delta x)) \\ &= \sum h(x)f(x)\Delta x \to \int h(x)f(x)dx.\end{aligned}$$

Data, long run average



$f(x) \to x_1, ..., x_i, ..., x_n$.
$n(x)$ = number of points in $(x, x + \Delta x)$.

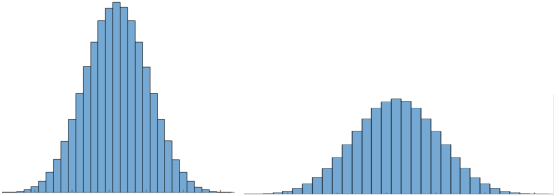
$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{x} xn(x) = \sum_{x} x\frac{n(x)}{n} \\ &\to \sum xP(X \in (x, x + \Delta x)) \\ &= \sum xf(x)\Delta x \to \int xf(x)dx = \mathbb{E}(X).\end{aligned}$$

Same logic for $\mathbb{E}(h(X))$.

Variance

Continuous

$$P(X \in (x, x+\Delta x)) = f(x)\Delta x.$$



$$\mathbb{E}(X) = \int x f(x) dx = \mu.$$

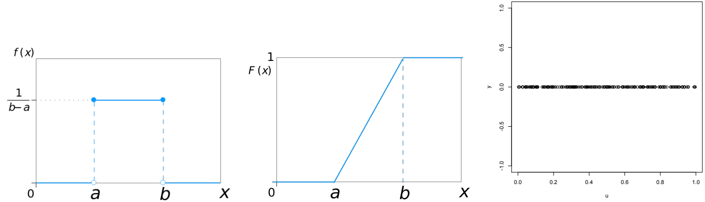
$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \int (x-\mu)^2 f(x) dx.$$

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

Fluctuation, volatility, spread.

Uniform

$U \sim \text{Uniform}[0,1]$, i.e., the density of U is $f(u) = 1$ for $u \in [0,1]$ (or $f(u) = 1/(b-a)$ if $u \in [a,b]$), $f(u) = 0$ otherwise.



$P(U \in (u, u+\Delta u)) = f(u)\Delta u = \Delta u.$

Imagine 1 million points distributed uniformly in [0, 1].

Number of points in $(u, u+\Delta u)$ is Δu million.

e.g., Number of points in $(.3, .31)$ is $.01$ million.

Uniform

$$F(u) = P(U \leq u) = \begin{cases} 0 & 0 < u \\ u & 0 \leq u \leq 1 \\ 1 & u > 1 \end{cases}$$

$F(u)$: proportion of points below u.

$$\mathbb{E}(U) = \int_0^1 u f(u) du = \frac{1}{2}.$$

$$\mathbb{E}(U^2) = \int_0^1 u^2 f(u) du = \frac{1}{3}.$$

$$\text{Var}(U) = \mathbb{E}(U^2) - (\mathbb{E}(U))^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

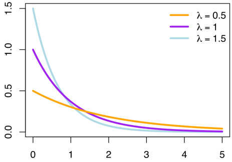
Pseudo-random number generator

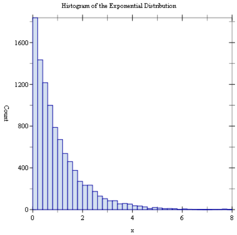
Linear congruential method
Start from an integer X_0, and iterate

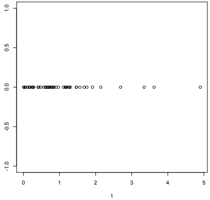
$$X_{t+1} = aX_t + b \bmod M.$$

Output $U_t = X_t/M$. e.g., $a = 7^5$, $b = 0$, and $M = 2^{31} - 1$.
mod: divide and take the remainder, e.g., $7 = 2 \bmod 5$.
e.g., $a = 7$, $b = 1$, $M = 5$, $X_0 = 1$, then
$X_1 = 1 \times 7 + 1 \bmod 5 = 3$.
$X_2 = 3 \times 7 + 1 \bmod 5 = 2$.

Exponential

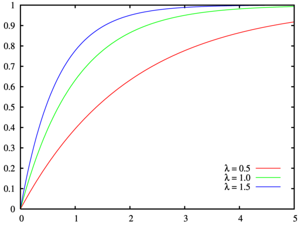


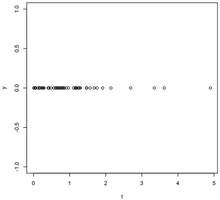




$T \sim \text{Exponential}(\lambda)$,
$f(t) = \lambda e^{-\lambda t}$ for $t \geq 0$,
$f(t) = 0$ for $t < 0$.
$P(T \in (t, t + \Delta t)) = \lambda e^{-\lambda t} \Delta t$.
Imagine 1 million particles, mark the times when they decay.
1 million points on real line. Their distribution is exponential.
Number of points in $(t, t + \Delta t)$ is $\lambda e^{-\lambda t} \Delta t$ million.

Exponential





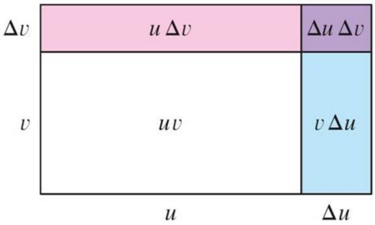
$$F(t) = \int_0^t f(t)dt = \int_0^t \lambda e^{-\lambda t} dt$$
$$= -e^{-\lambda t}|_0^t = 1 - e^{-\lambda t}.$$

$F(t)$: proportion of points below t
Half-life: $F(t_{\text{half}}) = P(T \le t_{\text{half}}) = 1/2$.
1 million particles, by half life, half million will have decayed.

Exponential expectation

$$
\begin{aligned}
\mathbb{E}(T) &= \int_0^\infty t\lambda e^{-\lambda t} dt \\
&= -\int_0^\infty t de^{-\lambda t} \\
&= -(te^{-\lambda t}|_0^\infty - \int_0^\infty e^{-\lambda t} dt) \\
&= -(0 - 0 + \frac{1}{\lambda} e^{-\lambda t}|_0^\infty) = \frac{1}{\lambda}.
\end{aligned}
$$

Integral by parts



$$\frac{d}{dx}u(x)v(x) = u'(x)v(x) + u(x)v'(x).$$

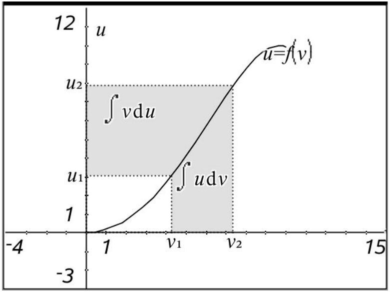
$$duv = udv + vdu.$$

$$\int [u'(x)v(x) + u(x)v'(x)]dx = u(x)v(x).$$

$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx.$$

$$\int udv = uv - \int vdu.$$

Integral by parts



$$\int u dv = uv - \int v du.$$

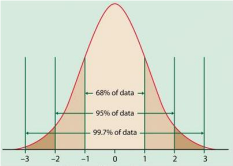
$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx.$$

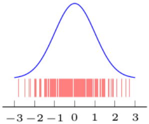
$$\frac{du(x)}{dx} = \frac{d}{dx}u(x) = u'(x);\ du(x) = u'(x)dx.$$

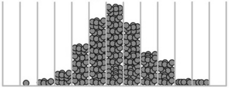
Normal or Gaussian

Let $Z \sim \mathrm{N}(0, 1)$, i.e., the density of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$







$$\int_{-2}^{2} f(z)dz = 95\%.$$

Normal expectation

100A

Ying Nian Wu

Discrete

Continuous

Let $Z \sim \mathrm{N}(0, 1)$, i.e., the density of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

$$\begin{aligned} \mathbb{E}(Z) &= \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= -\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} |_{-\infty}^{\infty} \\ &= 0. \end{aligned}$$

The density is symmetric around 0.

Normal variance

Let $Z \sim \mathrm{N}(0,1)$, i.e., the density of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

$$\begin{aligned}
\mathbb{E}(Z^2) &= \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-z) de^{-\frac{z^2}{2}} \\
&= \frac{1}{\sqrt{2\pi}} (-ze^{-\frac{z^2}{2}} |_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} d(-z)) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1.
\end{aligned}$$

$$\mathrm{Var}(Z) = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2 = 1.$$

Variance

For $X \sim f(x)$, let $\mu = \mathbb{E}(X)$.

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

$$
\begin{aligned}
\mathbb{E}[r(X) + s(X)] &= \int [r(x) + s(x)] f(x) dx \\
&= \int r(x) f(x) dx + \int s(x) f(x) dx \\
&= \mathbb{E}[r(X)] + \mathbb{E}[s(X)].
\end{aligned}
$$

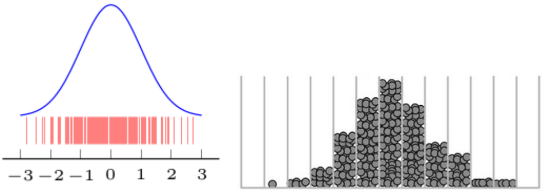
Linear transformation

For $X \sim f(x)$. Let $Y = aX + b$.

$$\begin{aligned}
\mathbb{E}(Y) = \mathbb{E}(aX+b) &= \int (ax+b)f(x)dx \\
&= a\int xf(x)dx + b\int f(x)dx \\
&= a\mathbb{E}(X) + b.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(Y) = \mathrm{Var}(aX+b) &= \mathbb{E}[((aX+b) - \mathbb{E}(aX+b))^2] \\
&= \mathbb{E}[(aX+b-(a\mathbb{E}(X)+b))^2] \\
&= \mathbb{E}[a^2(X-\mathbb{E}(X))^2] \\
&= a^2\mathbb{E}[(X-\mathbb{E}(X))^2] = a^2\mathrm{Var}(X).
\end{aligned}$$

Data



Sampling $f(x) \to x_1, ..., x_i, ..., x_n$
(e.g., random number generator → .22, .31, .92, .45, ...)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \to \mathbb{E}(X) = \mu.$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \to \mathrm{Var}(X) = \sigma^2$$

Data

Sampling $f(x) \to x_1, ..., x_i, ..., x_n$
(e.g., random number generator $\to$ .22, .31, .92, .45, ...)

$$y_i = ax_i + b.$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \to \mathbb{E}(X) = \mu.$$

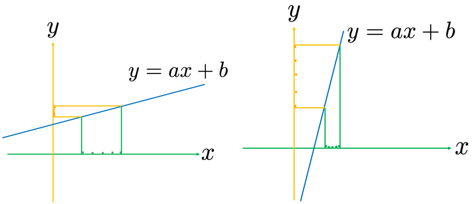
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = a\frac{1}{n}\sum_{i=1}^{n} x_i + b = a\bar{x} + b.$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - (a\bar{x} + b))^2 = \frac{1}{n}\sum_{i=1}^{n} a^2(x_i - \bar{x})^2.$$

Change of density under linear transformation

Change of variable

$X \sim f(x)$, $Y = aX + b$ $(a > 0)$. $Y \sim g(y)$.



$$y = ax + b,\ x = (y - b)/a.$$

$$P(X \in (x, x + \Delta x)) = P(Y \in (y, y + \Delta y)).$$

$$f(x)\Delta x = g(y)\Delta y.$$

$$g(y) = f(x)\frac{\Delta x}{\Delta y} = f((y - b)/a))/a.$$

Space warping, stretching or squeezing.

Normal or Gaussian

Let $Z \sim \mathrm{N}(0,1)$, i.e., the density of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Let $X = \mu + \sigma Z$. $Z = (X - \mu)/\sigma$. Then

$$\mathbb{E}(X) = \mathbb{E}(\mu + \sigma Z) = \mu + \sigma \mathbb{E}(Z) = \mu.$$

$$\mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2.$$

$$f(z)\Delta z = g(x)\Delta x.$$

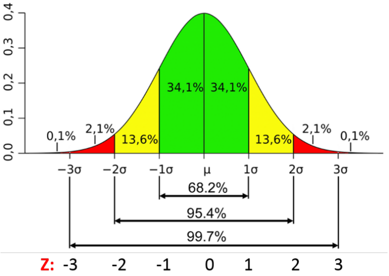
$$\begin{aligned} g(x) &= f(z)\frac{\Delta z}{\Delta x} \\ &= f((x-\mu)/\sigma)/\sigma \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \end{aligned}$$

Normal or Gaussian

Let $Z \sim \mathrm{N}(0,1)$. Let $X = \mu + \sigma Z$. $Z = (X-\mu)/\sigma$.
$X \sim \mathrm{N}(\mu, \sigma^2)$,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

(we now use $f(x)$ to denote the density of X.)



$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = P(-2 \le Z \le 2) = 95\%.$$