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

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 variables

Roll a die

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

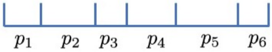
$$p(x) = P(X = x).$$

Capital letter: random variable
Lower case: particular value, running variable

$$X \sim p(x).$$

Probability distribution

Biased die:



Randomly throw a point into [0, 1], which bin (1, 2, ..., 6) it falls into?
$\omega \in \Omega = [0, 1]$, equally likely.
$X(\omega)$ is the bin that ω belongs to, not necessarily equally likely.
Throw 1 million points, what is the proportion of points in each bin? Or how often the points fall into each bin?

Probability distribution

Biased die:

x	1	2	3	4	5	6
p(x)	0.1	0.2	0.1	0.2	0.1	0.3
	10%	*20%*	*10%*	*20%*	*10%*	*30%*

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

$p(x)$: how often $X = x$.
$p(x)$: probability mass function, probability distribution, law
$\sum_x p(x) = 1$.
$P(X \in \{5, 6\}) = p(5) + p(6)$.
$P(X \in [a, b]) = \sum_{x \in [a,b]} p(x)$.

Expectation

Biased die

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%*

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

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

Biased die

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

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

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

Long run average of squared deviation from the mean.

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

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

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

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

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

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

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

Linear transformation

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

$$Y = aX + b.$$

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

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, ...)

$$y_i = ax_i + b.$$

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

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

Linear transformation

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

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

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

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

Data

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

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

$$y_i = a x_i + b.$$

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

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

Short-cut for variance

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

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

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

Transformation

$$h(x) = ax + b.$$

$$\mathbb{E}[h(X)] = \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = h(\mathbb{E}(X)).$$

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

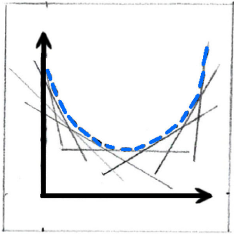
$$h(x) = x^2.$$

$$\mathbb{E}[h(X)] = \mathbb{E}(X^2);$$

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

Convex function

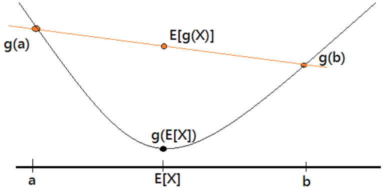
Upper envelop and supporting lines



$$g(x) \geq a_0 x + b_0;\; g(x_0) = a_0 x_0 + b_0.$$

Supporting line at x_0 touches $g(x)$ at x_0, but below $g(x)$ at other places.

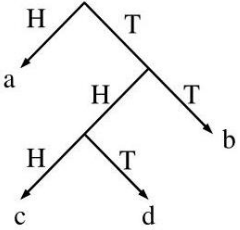
$P(X = a) = P(X = b) = 1/2.$
$\mathbb{E}(X) = (a + b)/2,\ g(\mathbb{E}(X)) = g((a + b/2).$
$\mathbb{E}(g(X)) = (g(a) + g(b))/2.$
$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)).$



$x_0 = \mathbb{E}(X).$
$g(x_0) = a_0 x_0 + b_0$ (supporting line at x_0)
$g(x) \geq a_0 x + b_0.$
$\mathbb{E}(g(X)) \geq \mathbb{E}(a_0 X + b_0) = a_0 \mathbb{E}(X) + b_0 = a_0 x_0 + b_0 = g(\mathbb{E}(X)).$

Entropy

	a	b	c	d
Pr	1/2	1/4	1/8	1/8
flips	*H*	*TT*	*THH*	*THT*



Coin flippings

Ω	a	b	c	d
Pr	1/2	1/4	1/8	1/8
$-\log p$	1	2	3	3

$$H(X) = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = \frac{7}{4}\textbf{flips}$$

	a	b	c	d
Pr	1/2	1/4	1/8	1/8
flips	H	TT	THH	THT

Prefix code

	a	b	c	d
Pr	1/2	1/4	1/8	1/8
code	1	00	011	010

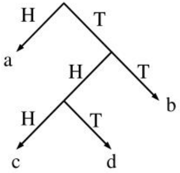
100101100010→abacbd

$$\mathbf{E}[l(X)] = \sum_x l(x)p(x) = 1\times\frac{1}{2} + 2\times\frac{1}{4} + 3\times\frac{1}{8} + 3\times\frac{1}{8} = \frac{7}{4}\,\textbf{bits}$$

Coding

Optimal code

	a	b	c	d
Pr	1/2	1/4	1/8	1/8
code	1	00	011	010



100101100010→abacbd

Sequence of coin flipping
A completely random sequence
Cannot be further compressed

$$l(x) = -\log p(x)$$

$$\mathbf{E}[l(X)] = H(p)$$

e.g., two words I, probability

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 exactly k heads.
e.g., $n = 3$,
$P(X = 2) = P(HHT) + P(HTH) + P(THH) = 3p^2(1-p)$.
$p = 1/2$, we have $P(X = k) = \binom{n}{k}/2^n$.

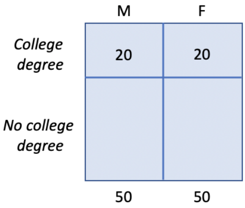
Recall independence

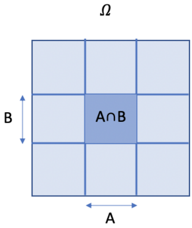
Definition 1:

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

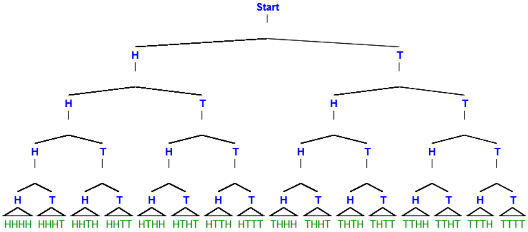
Definition 2:

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





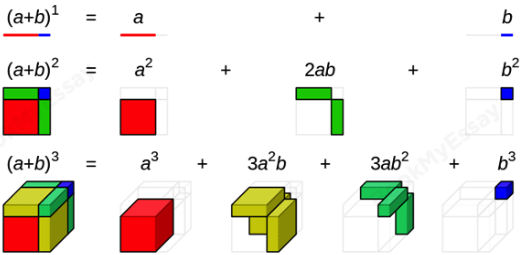
Binomial formula



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

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

Binomial



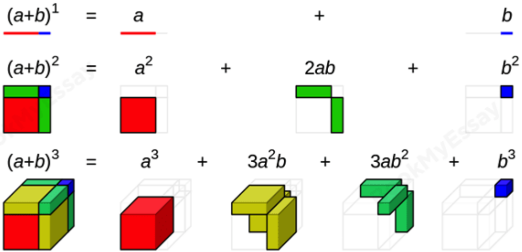
$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$$

Let $a = p$, $b = q = 1 - p$. Randomly throw a point into unit cube, equally likely setting.

Each rectangular piece corresponds to a particular sequence.

Each color corresponds to a particular number of heads.

Binomial



Let $a = p$, $b = q = 1 - p$.
$n = 2$, $P(X = 2) = P(HH) = p^2$.
$P(X = 0) = P(TT) = (1 - p)^2$.
$P(X = 1) = P(HT) + P(TH) = 2p(1 - p)$.
$n = 3$, $P(X = 3) = P(HHH) = p^3$.
$P(X = 2) = P(HHT) + P(HTH) + P(THH) = 3p^2(1 - p)$.

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.

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

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

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

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

Law of large number
Probability = long run frequency

Law of large number

long run frequency $\to$ probability
Flip a fair coin independently $\to 2^n$ sequences, Ω.

$$A_\epsilon = \{\omega : X(\omega)/n \in (1/2-\epsilon, 1/2+\epsilon)\},$$

the set of sequences whose frequencies of heads are close to 1/2.

$$P(X/n \in (1/2-\epsilon, 1/2+\epsilon)) = \frac{|A_\epsilon|}{|\Omega|} \to 1.$$

Almost all the sequences have frequencies of heads close to 1/2.
e.g., $n = 1$ million. Almost all the $2^{1 \text{ million}}$ sequences have frequencies of heads to be within [.49, .51].

Law of large number

e.g., $n = 1$ million. Almost all the $2^{1 \text{ million}}$ sequences have frequencies of heads to be within [.49, .51].

$$P(X/1m \in [.49, .51]) = P(X \in [.49m, .51m])$$
$$= \sum_{k=.49m}^{.51m} \binom{1m}{k} / 2^{1m} \approx 1.$$

Binomial expectation

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

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

Binomial variance

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

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

Binomial variance

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

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

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

Survey sampling

A box with R red balls and B blue balls. $N = R + B$ balls in total.
Randomly pick a ball. $P(red) = R/N = p$.
Randomly pick n balls sequentially (with replacement, put the picked ball back). Let $X =$ number of red balls.
The distribution of X:
$X \sim \text{Binomial}(n, p = R/N)$.
Survey sampling, poll.

Survey sampling

$\Omega =$ all N^n sequences, equally likely.
$X(\omega)$: number of red balls in sequence $\omega \in \Omega$.
$A_k = \{\omega : X(\omega) = k\}$, all sequences with k read balls.
Choose k blanks from n blanks. For each chosen blank, fill in a red ball. For each unchosen blank, fill in a blue ball.

$$|A_k| = \binom{n}{k} R^k B^{n-k}.$$

Axiom 0:

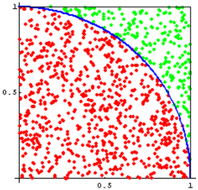
$$P(A_k) = P(X = k) = \frac{|A_k|}{|\Omega|} = \binom{n}{k} p^k (1-p)^{n-k}.$$

$\mathbb{E}(X) =$ average of $X(\omega)$ over all N^n sequences.
$\text{Var}(X) =$ variance of $X(\omega)$ over all N^n sequences.

Law of large number

Ω = all N^n sequences, equally likely.
$X(\omega)$: number of red balls in sequence $\omega \in \Omega$.
$X(\omega)/n$: frequency of red balls in sequence ω.
$\mathbb{E}(X/n) = p = R/N$ = average of $X(\omega)/n$ over all the N^n sequences in Ω.
$\text{Var}(X/n) = p(1-p)/n$ = variance of $X(\omega)/n$ over all the N^n sequences in Ω.
Law of large number: Among all N^n equally likely sequences, almost all of them have $X(\omega)/n$ close to p.

Randomly throw n points into the unit square. Let m be the number of points falling below the curve.



The distribution of m is:
$m \sim \text{Binomial}(n, p = \pi/4)$.
Ω = all possible sequences of points = $[0, 1]^{2n}$.

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

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

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

Geometric series

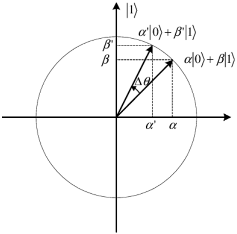
$$
\begin{aligned}
(1-a)(1+a+...+a^m) &= 1+a+...+a^m \\
&\quad -(a+a^2+...+a^m+a^{m+1}) \\
&= 1-a^{m+1}.
\end{aligned}
$$

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

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: $p(0) = |\alpha|^2$, $p(1) = |\beta|^2$.

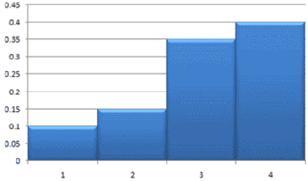
Schrodinger cat: P(alive) $= (1/\sqrt{2})^2 = 1/2$.

Recall discrete random variable

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

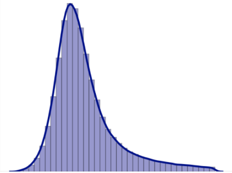
Probability histogram

x	P(x)
1	0.10
2	0.15
3	0.35
4	0.40



area of bin $x = p(x)$.
Randomly throw a point ω into the whole blue region Ω,
$X(\omega) = x$ if ω falls into bin x.

Continuous random variable



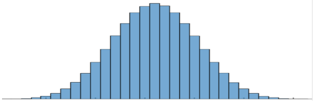
Continuous (e.g., height).
Discretize x-axis into equally spaced bins $(x, x + \Delta x)$, e.g., (6 ft, 6 ft 1 inch), precision = 1 inch.

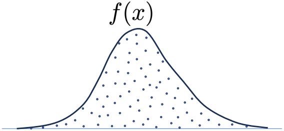
$$P(X \in (x, x + \Delta x)) = f(x)\Delta x,$$

$f(x)$: height of bin $(x, x + \Delta x)$. $f(x)\Delta x$: area.
Randomly throw a point ω into the whole blue region Ω, $X(\omega) = x$ if the point falls into bin $(x, x + \Delta x)$.

Probability density function

Let $\Delta x \to 0$, continuous.





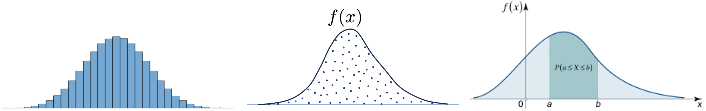
Randomly throw a point ω into the region Ω under curve $f(x)$.
Return $X(\omega)$ = horizontal coordinate of the point.
Repeat n (e.g., 1000) times, frequency $\to$ probability,

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

How often $X \in (x, x + \Delta x)$.

Probability density function

Let $\Delta x \to 0$, continuous.



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

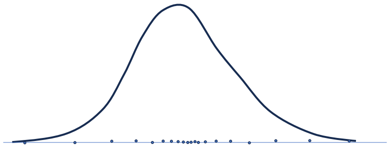
$$P(X \in (a, b)) \approx \sum_{\text{bins}} f(x)\Delta x \to \int_a^b f(x)dx.$$

area under $f(x)$ between a and b.

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Scatterplot

Collapse the points onto x-axis.

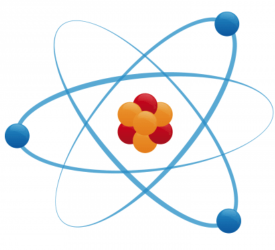


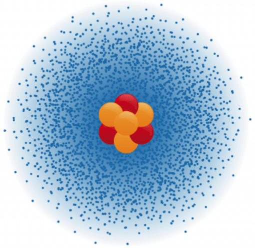
Sample density or distribution of the n points.
Sample density at x = number of points in $(x, x + \Delta x)/\Delta x$.
Normalize the count m to frequency m/n, as $n \to \infty$,
frequency $\to$ probability,

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

Point cloud

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

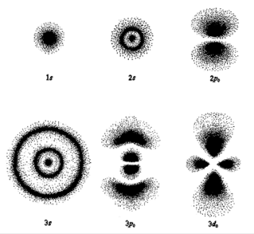


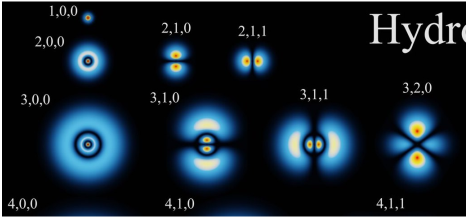


Prob density = prob mass in the cell / volume of cell.

Point cloud

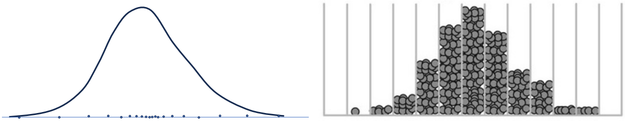
Electron cloud, heat map, prob density





Observer: $f(x) = |\psi(x)|^2$.
Prob density = prob mass in the cell / volume of cell.

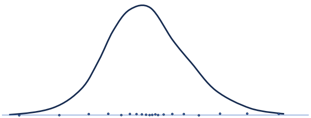
Sample

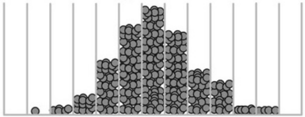


Sample density or distribution of the n points.
Sample histogram (can fluctuate if do it again).
Normalize the count to frequency $\rightarrow$ probability,

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

Population





Population density or distribution of the N (e.g., 300 million) points.

Population histogram (no fluctuation, always the same).

Normalize the count to proportion,

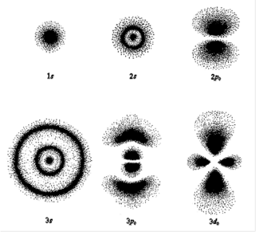
$$f(x) = \frac{P(X \in (x, x + \Delta x))}{\Delta x}.$$

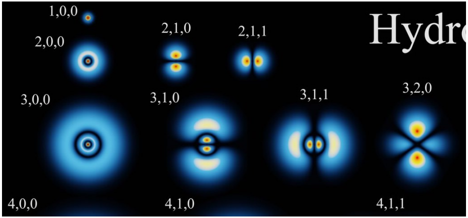
Population (300 million, fixed) $\to$ sample (1000, fluctuate)

Population $\to$ sample (1 million, fluctuation diminishes)

Population

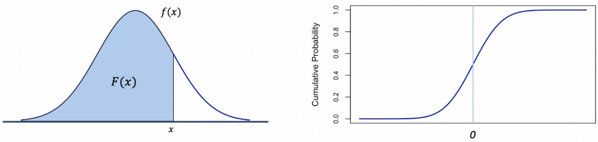
Electron cloud, heat map, prob density





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

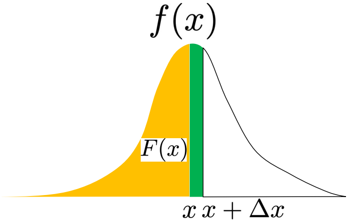
Cumulative density function

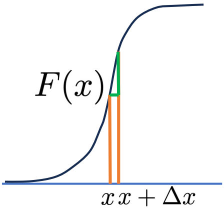


$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx.$$

SAT score $x \to$ 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:

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

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

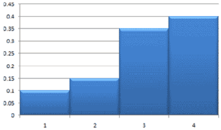
Expectation

Recall discrete

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

Probability histogram

x	P(x)
1	0.10
2	0.15
3	0.35
4	0.40



area of bin $x = p(x)$.

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

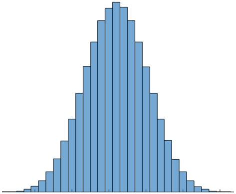
$$\mathbb{E}[h(X)] = \sum h(x)P(X = x) = \sum h(x)p(x).$$

Long run average.

Expectation

Continuous

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

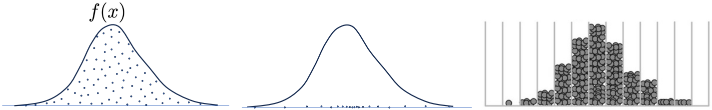


$$\mathbb{E}(X) = \sum xP(X \in (x, x+\Delta x)) = \sum xf(x)\Delta x \to \int xf(x)dx.$$

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

Long run average, center.

Data



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

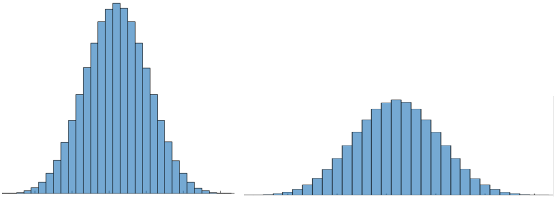
$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{\text{bins}} x m_x = \sum_{\text{bins}} x \frac{m_x}{n} \\
&\to \sum x P(X \in (x, x + \Delta x)) \\
&= \sum x f(x) \Delta x \to \int x f(x) dx = \mathbb{E}(X).
\end{aligned}
$$

Long run average. Same logic for $\mathbb{E}(h(X))$.
Same logic for population average.

Variance

Continuous

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



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

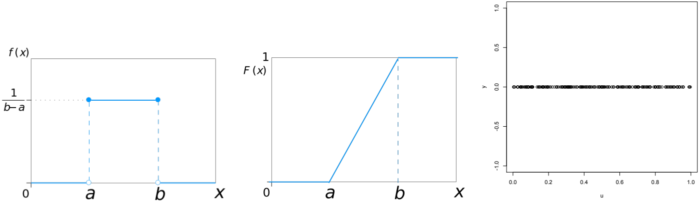
$$\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2] = \int (x - \mu)^2 f(x) dx.$$

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

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

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

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

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

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

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

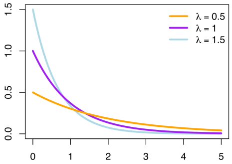
Pseudo-random number generator

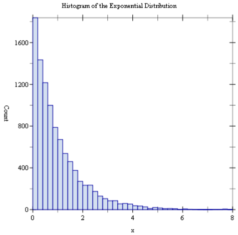
Linear congruential method
Start from an integer X_0, and iterate

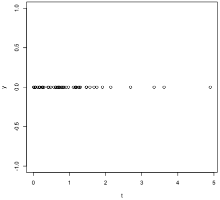
$$X_{t+1} = aX_t + b \bmod M.$$

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., $7 = 2$ mod 5.
e.g., $a = 7$, $b = 1$, $M = 5$, $X_0 = 1$, then
$X_1 = 1 \times 7 + 1$ mod $5 = 3$.
$X_2 = 3 \times 7 + 1$ mod $5 = 2$.

Exponential

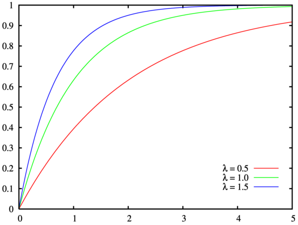


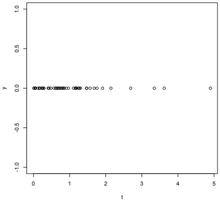




$T \sim \text{Exponential}(\lambda)$,
$f(t) = \lambda e^{-\lambda t}$ for $t \geq 0$,
$f(t) = 0$ for $t < 0$.
$P(T \in (t, t + \Delta t)) = \lambda e^{-\lambda t} \Delta t$.
Imagine 1 million particles, mark the times when they decay.
1 million points on real line. Their distribution is exponential.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

Exponential





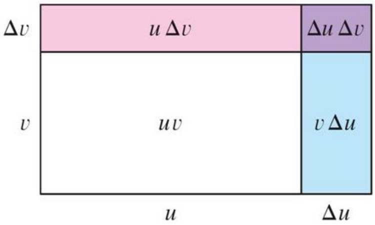
$$F(t) = \int_0^t f(t)dt = \int_0^t \lambda e^{-\lambda t} dt$$
$$= -e^{-\lambda t}|_0^t = 1 - e^{-\lambda t}.$$

$F(t)$: proportion of points below t
Half-life: $F(t_{\text{half}}) = P(T \leq t_{\text{half}}) = 1/2$.
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



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

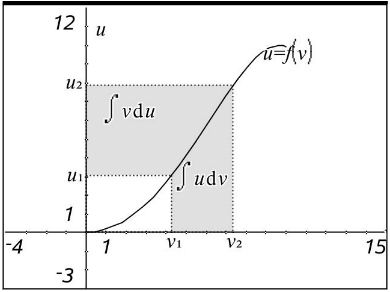
$$duv = udv + vdu.$$

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

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

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

Integral by parts



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

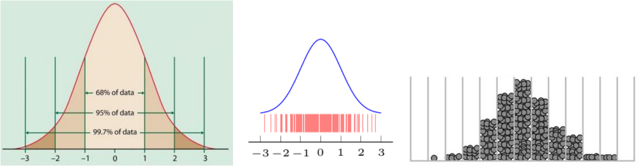
$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx.$$

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

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

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

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

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

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

Variance

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

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

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

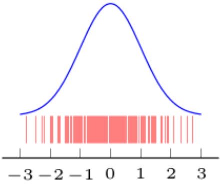
Linear transformation

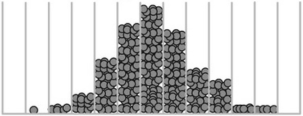
For $X \sim f(x)$. Let $Y = aX + b$.

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

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

Data





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

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

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

Data

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

$$y_i = ax_i + b.$$

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

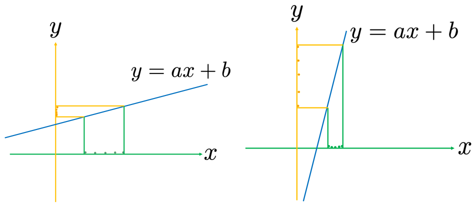
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = a\frac{1}{n}\sum_{i=1}^{n} x_i + b = a\bar{x} + b.$$

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

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

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

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

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

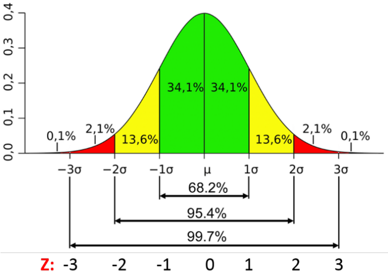
$$\begin{aligned} g(x) &= f(z)\frac{\Delta z}{\Delta x} \\ &= f((x-\mu)/\sigma)/\sigma \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \end{aligned}$$

Normal or Gaussian

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

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

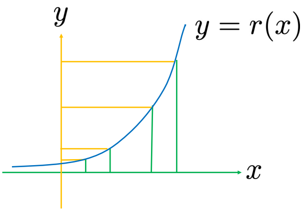
(we now use $f(x)$ to denote the density of X.)



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

Non-linear transformation

$X \sim f(x)$, $Y = r(X)$, monotone. $Y \sim g(y)$.



$$y = r(x),\ x = r^{-1}(y).$$

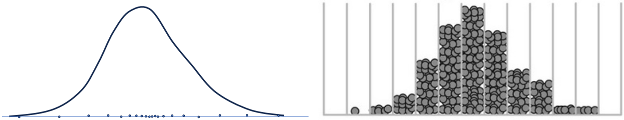
$$P(X \in (x, x+\Delta x)) = P(Y \in (y, y+\Delta y)).$$

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

$$\Delta y/\Delta x = r'(x).$$

Locally linear, space warping.

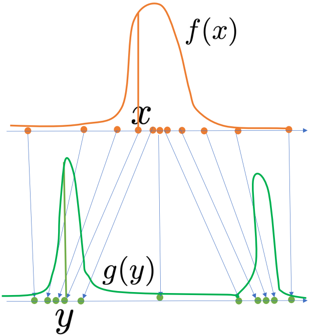
Space warping



Squeezing or stretching the bins → changes the density and histogram.

Non-linear transformation

$X \sim f(x)$, $Y = r(X)$, monotone. $Y \sim g(y)$.



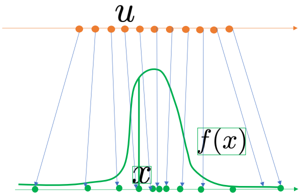
$$y = r(x),\ x = r^{-1}(y).$$

Order preserving mapping:

$$P(X \leq x) = P(Y \leq y).$$

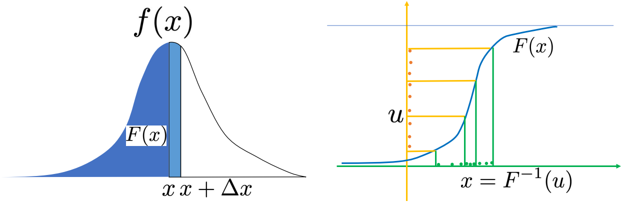
$$F(x) = G(y).$$

Inversion method



$U \sim \text{Unif}[0,1]$.
$P(U \leq u) = P(X \leq x)$.
$u = F(x)$, $x = F^{-1}(u)$.
Population: $\{x_1, x_2, ..., x_N\}$ (ordered).
Sample $i \sim \text{Uniform}\{1, 2, ..., N\}$, return x_i.
$P(X \leq x_i) = i/N = F(x_i)$.
$U = i/N \sim \text{Uniform}[0,1]$, $x_i = F^{-1}(U)$.

Inversion method



$U \sim \text{Unif}[0,1]$. $X = F^{-1}(U)$. Then $f(x) = F'(x)$ is the pdf of X.

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

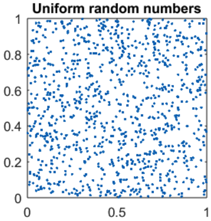
$$\Delta u = f(x)\Delta x.$$

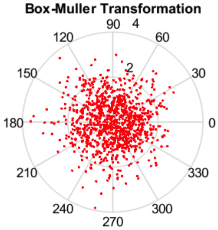
$$f(x) = \frac{\Delta u}{\Delta x} = F'(x).$$

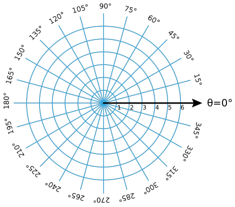
Inversion method

Suppose we want to generate $X \sim \text{Exponential}(1)$.
$F(x) = 1 - e^{-x}$.
$F(x) = u$, i.e., $1 - e^{-x} = u$, $e^{-x} = 1 - u$. $x = -\log(1 - u)$.
Generate $U \sim \text{Unif}[0, 1]$. Return $X = -\log(1 - U)$.

Polar method







Polar method

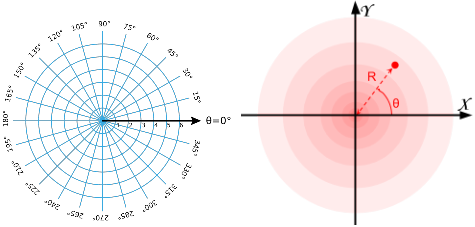
$X \sim \mathrm{N}(0, 1)$, $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.
$Y \sim \mathrm{N}(0, 1)$, $f(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$.
X and Y are independent.

$$\begin{aligned} & P(X \in (x, x + \Delta x), Y \in (y, y + \Delta y)) \\ = \; & P(X \in (x, x + \Delta x)) \times P(Y \in (y, y + \Delta y)). \end{aligned}$$

$$f(x, y)\Delta x \Delta y = f(x)\Delta x \times f(y)\Delta y.$$

$$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right).$$

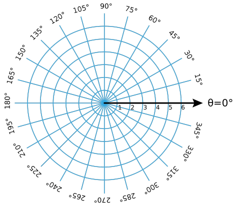
Polar method

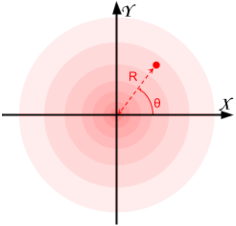


$x = r\cos\theta$, $y = r\sin\theta$.
Area of ring $R \in (r, r+\Delta r)) = 2\pi r\Delta r$.
Count proportion of points in the ring = density × area.

$$P(R \in (r, r+\Delta r)) = \frac{1}{2\pi}\exp\left(-\frac{r^2}{2}\right)2\pi r\Delta r$$
$$= \exp\left(-\frac{r^2}{2}\right)r\Delta r = \exp\left(-\frac{r^2}{2}\right)d\frac{r^2}{2}.$$

Polar method





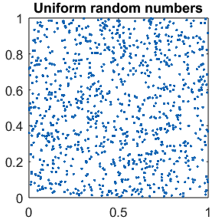
$x = r\cos\theta$, $y = r\sin\theta$.
Let $t = r^2/2$. $\Delta t = r\Delta r$.

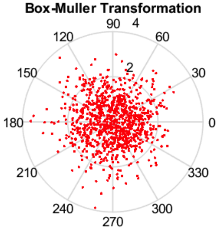
$$P(T \in (t, t+\Delta t)) = P(R \in (r, r+\Delta r)).$$

$$f(t)\Delta t = \exp\left(-\frac{r^2}{2}\right) r\Delta r = \exp(-t)\Delta t.$$

$$T \sim \text{Exponential}(1).$$

Polar method



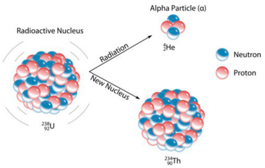


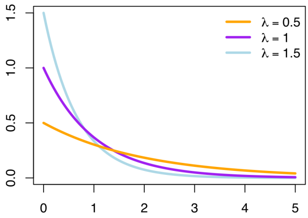
$T = -\log(1 - U_1)$.
$R = \sqrt{2T}$.
$\theta = 2\pi U_2$.
$X = R\cos\theta$, $Y = R\sin\theta$.
$(U_1, U_2) \to (X, Y)$.

Non-linear transformation

$X \sim f(x)$, $Y = r(X)$. $Y \sim g(y)$.
X consists of iid Gaussian $\mathrm{N}(0, 1)$ noises.
r is learned from training examples by neural network (deep learning).

Particle decay





T: time until decay.
$T \sim \text{Exponential}(\lambda)$.
$P(T \in (t, t + \Delta t)) = f(t)\Delta t = \lambda e^{-\lambda t}\Delta t$.

Making a movie

Divide the time into small intervals of length Δt (e.g., 1/24 second, or 1/100 second).

Show a picture at 0, Δt, $2\Delta t$, ...

Give an illusion of continuous time process as $\Delta t \to 0$.

Continuous time process

Bank account

Divide $[0, t]$ into n small intervals, $\Delta t = t/n$.
Interest rate $= r$.
Time 0: $\$1$.
Time Δt: $\$(1 + r\Delta t)$.
Time $2\Delta t$: $\$(1 + r\Delta t)^2$.
Time $3\Delta t$: $\$(1 + r\Delta t)^3$.
...
Time $t = n\Delta t$: $\$(1 + r\Delta t)^n$.

$$\left(1 + r\frac{t}{n}\right)^n \to e^{rt},$$

as $n \to \infty$ or $\Delta t \to 0$.

Continuous time process

Bank account

Divide $[0, t]$ into n small intervals, $\Delta t = t/n$.
Interest rate $= r$.

$$\left(1 + \frac{1}{n}\right)^n \to e.$$

$$1 + \frac{1}{n} \doteq e^{1/n}.$$

$$1 + \Delta x \doteq e^{\Delta x}.$$

$$\left(1 + r\frac{t}{n}\right)^n \to e^{rt}.$$

$$(1 + r\Delta t)^{t/\Delta t} \doteq \left(e^{r\Delta t}\right)^{t/\Delta t} = e^{rt}.$$

Poisson process

Flip a coin within each interval.
$p = \lambda \Delta t$ (e.g., $\Delta t = 1$ hour. $\lambda =$ once every 10 year. $\lambda \Delta t = 1/3650 \times 1/24$).

Geometric waiting time

$$\begin{aligned} P(T \in (t, t+\Delta t)) &= (1-\lambda \Delta t)^{t/\Delta t} \lambda \Delta t \\ &\doteq \left(e^{-\lambda \Delta t}\right)^{t/\Delta t} \lambda \Delta t = e^{-\lambda t} \lambda \Delta t. \end{aligned}$$

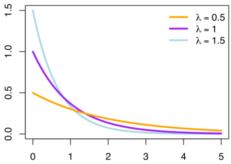
Exponential distribution

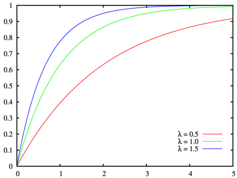
Flip a coin within each interval.
$p = \lambda \Delta t$ (e.g., $\Delta t = .001$ second. $\lambda =$ once every minute. $\lambda \Delta t = 1/60 \times .001$).

Exponential waiting time

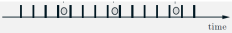
$$\frac{P(T \in (t, t+\Delta t))}{\Delta t} = \lambda e^{-\lambda t}.$$

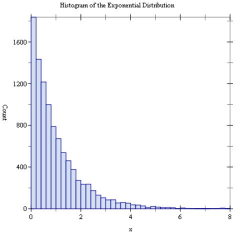
$$P(T > t) = (1 - \lambda \Delta t)^{t/\Delta t} \doteq (e^{-\lambda \Delta t})^{t/\Delta t} = e^{-\lambda t}.$$





Exponential = geometric





1 million particles decay in different period. Each small period is a bin.

Geometric waiting time

We can write $T = \tilde{T}\Delta t$, where $\tilde{T} \sim \text{Geometric}(p = \lambda \Delta t)$. Then

$$\mathbb{E}(T) = \mathbb{E}(\tilde{T})\Delta t = \frac{1}{p}\Delta t = \frac{1}{\lambda \Delta t}\Delta t = 1/\lambda.$$

Poisson distribution

Flip a coin within each interval.
Let X be the number of heads within $[0, t]$, then
$X \sim \text{Binomial}(n = t/\Delta t, p = \lambda \Delta t)$.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \to \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

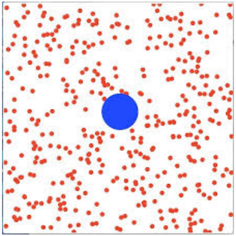
$\mathbb{E}(X) = np = (t/\Delta t)(\lambda \Delta t) = \lambda t$.
$\lambda = \mathbb{E}(X)/t$, rate or intensity.

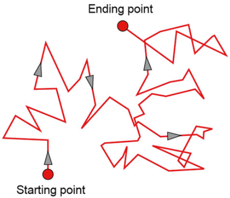
Poisson distribution

$$
\begin{aligned}
P(X = k) &= \frac{n(n-1)...(n-k+1)}{k!} p^k (1-p)^{n-k} \\
&= \frac{t/\Delta t(t/\Delta t - 1)...(t/\Delta t - k + 1)}{k!} \\
&\times (\lambda \Delta t)^k (1 - \lambda \Delta t)^{t/\Delta t - k} \\
&= \frac{t(t-\Delta t)(t - 2\Delta t)...(t-(k-1)\Delta t)}{k!} \\
&\times \lambda^k (1-\lambda \Delta t)^{t/\Delta t} (1 - \lambda \Delta t)^{-k} \\
&\to \frac{t^k}{k!} \lambda^k (e^{-\lambda \Delta t})^{t/\Delta t} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.
\end{aligned}
$$

Diffusion or Brownian motion

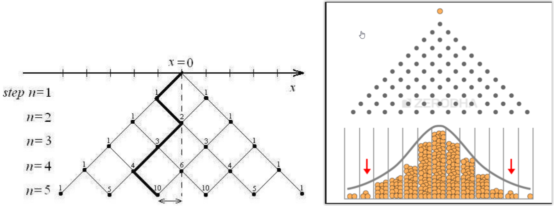
Dust particle in water





Recall random walk

Either go forward or backward by flipping a fair coin.



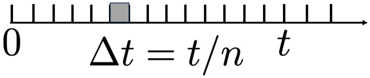
Number of heads $Y \sim \text{Binomial}(n, 1/2)$, then random walk ends up at X,

$$X = Y - (n - Y) = 2Y - n.$$

$$X = \epsilon_1 + \epsilon_2 + ... + \epsilon_n.$$

$\epsilon_k = 1$ or -1 with probability 1/2 each.

Discretize time and space

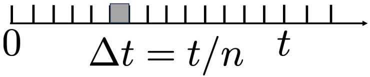


(1) Time: Divide $[0, t]$ into n intervals, $\Delta t = t/n$ (time unit).
(2) Space: Within each small time interval, move forward or backward by Δx (space unit).
$P(\epsilon_i = 1) = P(\epsilon_i = -1) = 1/2$. ϵ_i are independent.

$$X = \sum_{i=1}^{n} \epsilon_i \Delta x = (Y - (n - Y))\Delta x = (2Y - n)\Delta x.$$

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(\epsilon_i)\Delta x = \mathbb{E}(2Y - n)\Delta x = 0.$$

Diffusion or Brownian motion



$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(\epsilon_i)\Delta x^2 = n\Delta x^2 = \frac{t}{\Delta t}\Delta x^2.$$

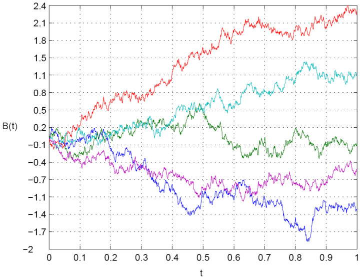
$$\text{Var}(X) = \text{Var}((2Y - n)\Delta x) = 4\text{Var}(Y)\Delta x^2 = n\Delta x^2.$$

$$\Delta x^2/\Delta t = \sigma^2;\ \Delta x = \sigma\sqrt{\Delta t};\ \text{Var}(X) = \sigma^2 t.$$

$$\text{velocity} = \Delta x/\Delta t = \sigma/\sqrt{\Delta t} \to \infty.$$

Einstein, σ related to the size of molecules.

Diffusion or Brownian motion



$X = B(t)$.
Nowhere differentiable.
σ: volatility of stock price, basis for option pricing.
A drop of milk (millions of particles) diffuses in coffee.

Normal approximation

Central limit theorem

$P(\epsilon_i = 1) = P(\epsilon_i = -1) = 1/2$. ϵ_i are independent.

$$X = \sum_{i=1}^{n} \epsilon_i \Delta x = (2Y - n)\Delta x \sim \mathrm{N}(0, \sigma^2 t),$$

as $n \to 0$.

Sum of independent random variables $\sim$ Normal distribution.

Normal approximation

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

$$Z = \frac{X - \mu}{\sigma} = \frac{X - n/2}{\sqrt{n}/2},$$

then $\mathbb{E}(Z) = 0$, $\text{Var}(Z) = 1$, no matter what n is.
Z takes discrete values, with spacing $\Delta z = 1/\sigma = 2/\sqrt{n}$.

$$P(Z \in (a, b)) = \sum_{z \in (a,b)} p(z) \doteq \sum_{z \in (a,b)} f(z)\Delta z \to \int_a^b f(z)dz,$$

where $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is the density of $\text{N}(0, 1)$.

$$p(z)/\Delta z \to f(z).$$

Proof

Step 1:

$$p(0) \doteq \frac{1}{\sqrt{2\pi}} \Delta z.$$

Step 2:

$$\frac{p(z)}{p(0)} \doteq e^{-z^2/2}.$$

$X = \mu + Z\sigma = n/2 + Z\sqrt{n}/2.$

$$p(0) = P(X = n/2).$$

$$\frac{p(z)}{p(0)} = \frac{P(X = n/2 + z\sqrt{n}/2)}{P(X = n/2)} = \frac{P(X = n/2 + d)}{P(X = n/2)}.$$

Proof

$$P(X = k) = \frac{\binom{n}{k}}{2^n} = \frac{n!}{k!(n-k)!2^n},$$

For big n,

$$n! \sim \sqrt{2\pi n} n^n e^{-n},$$

$$\begin{aligned} P(X = n/2) &\sim \frac{n!}{(n/2)!^2 2^n} \\ &\sim \frac{\sqrt{2\pi n} n^n e^{-n}}{(\sqrt{2\pi(n/2)}(n/2)^{n/2})^2 2^n} \\ &\sim \frac{1}{\sqrt{2\pi}} \frac{2}{\sqrt{n}}. \end{aligned}$$

Proof

Let $k = \mu + z\sigma = n/2 + z\sqrt{n}/2 = n/2 + d$.

$$\frac{P(X = n/2 + d)}{P(X = n/2)} = \frac{\binom{n}{n/2+d}}{\binom{n}{n/2}}$$

$$= \frac{n!/[(n/2+d)!(n/2-d)!]}{n!/[(n/2)!(n/2)!]}$$

$$= \frac{(n/2)!(n/2)!}{(n/2+d)!(n/2-d)!}$$

$$= \frac{(n/2)(n/2-1)...(n/2-(d-1))}{(n/2+1)(n/2+2)...(n/2+d)}$$

$$= \frac{1(1-2/n)(1-2\times 2/n)...(1-(d-1)\times 2/n)}{(1+2/n)(1+2\times 2/n)...(1+d\times 2/n)}$$

$$= \frac{(1-\delta)(1-2\delta)...(1-(d-1)\delta)}{(1+\delta)(1+2\delta)...(1+d\delta)}$$

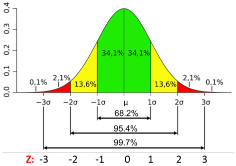
Proof

$$
\begin{aligned}
\rightarrow\ & \frac{e^{-\delta}e^{-2\delta}...e^{-(d-1)\delta}}{e^{\delta}e^{2\delta}...e^{d\delta}} \\
=\ & \frac{e^{-(1+2+...+(d-1))\delta}}{e^{(1+2+...+d)\delta}} \\
=\ & \frac{e^{-d(d-1)\delta/2}}{e^{d(d+1)\delta/2}} \\
=\ & e^{-[d(d-1)/2+d(d+1)/2]\delta} = e^{-d^2\delta} \\
=\ & e^{-(z\sqrt{n}/2)^2(2/n)} = e^{-\frac{z^2}{2}},
\end{aligned}
$$

where $\delta = 2/n$, and $d = z\sqrt{n}/2$.

Normal approximation

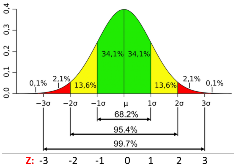
Let $X \sim \mathrm{Binomial}(n, p)$, sum of independent Bernoulli.
$\mathbb{E}(X) = np$, $\mathrm{Var}(X) = np(1-p)$.
$\mathbb{E}(X/n) = p$, $\mathrm{Var}(X/n) = p(1-p)/n$.
Approximately,
$X \sim \mathrm{N}(np, np(1-p))$.
$X/n \sim \mathrm{N}(p, p(1-p)/n)$.
e.g., $n = 100$, $p = 1/2$. $X \sim \mathrm{N}(50, 25)$.
$P(X \in [50 - 2 \times 5, 50 + 2 \times 5]) = P(X \in [40, 60]) = 95\%$.



Recall $\sum_{k=40}^{60} \binom{100}{k} / 2^{100} \to$ integral.

Normal approximation

Let $X \sim \text{Binomial}(n, p)$, sum of independent Bernoulli.
$\mathbb{E}(X) = np$, $\text{Var}(X) = np(1-p)$.
$\mathbb{E}(X/n) = p$, $\text{Var}(X/n) = p(1-p)/n$.
Approximately,
$X \sim \text{N}(np, np(1-p))$.
$X/n \sim \text{N}(p, p(1-p)/n)$.
e.g., Polling $n = 100$, $p = .2$. $X/n \sim \text{N}(.2, .04^2)$.
$P(X/n \in [.2 - 2 \times .04, .2 + 2 \times .04]) = P(X/n \in [.12, .28]) = 95\%$.



Normal approximation

Let $X \sim \text{Binomial}(n, p)$, sum of independent Bernoulli.
$\mathbb{E}(X) = np$, $\text{Var}(X) = np(1-p)$.
$\mathbb{E}(X/n) = p$, $\text{Var}(X/n) = p(1-p)/n$.
Approximately,
$X \sim \text{N}(np, np(1-p))$.
$X/n \sim \text{N}(p, p(1-p)/n)$.
e.g., Monte Carlo $n = 10000$, $p = \pi/4$.
$4m/n \sim \text{N}(\pi, \pi(4-\pi)/10000)$.