Probability and Statistics

Prof. Asim Tewari
IIT Bombay

Monty Hall Problem Puzzle

- The host of a game show, offers the guest a choice of three doors. Behind one is a expensive car, but behind the other two are goats.
  After you have chosen one door, he reveals one of the other two doors behind which is a goat (he wouldn't reveal a car).

  Now he gives you the chance to switch to the other unrevealed door or stay at your initial choice. You will then get what is behind that door.

  You cannot hear the goats from behind the doors, or in any way know which door has the prize.

  Should you stay, or switch, or doesn't it matter?

-

  Your first choice has a 1/3 chance of having the car, and that does not change.
  The other two doors HAD a combined chance of 2/3, but now a Goat has been revealed behind one, all the 2/3 chance is with the other door.

Probability Of Second Girl Child

- A family has two kids, one of them is a girl. Assume safely that the probability of each gender is 1/2. What is the probability that the other kid is also a girl?

- ANS:

1/3

This is a famous question in understanding conditional probability, which simply means that given some information you might be able to get a better estimate.

The following are possible combinations of two children that form a sample space in any earthly family:
Girl – Girl
Girl – Boy
Boy – Girl
Boy - Boy

Since we know one of the children is a girl, we will drop the Boy-Boy possibility from the sample space.
This leaves only three possibilities, one of which is two girls. Hence the probability is 1/3

Hard Logic Probabilty Puzzle

- Bruna was first to arrive at a 100 seat theater.
  She forgot her seat number and picks a random seat for herself.

  After this, every single person who get to the theater sits on his/her seat if its available else chooses any available seat at random.
  Neymar is last to enter the theater and 99 seats were occupied.

  Whats the probability what Neymar gets to sit in his own seat ?

- ANS: 1/2

  one of two is the possibility
  1. If any of the first 99 people sit in neymar seat, neymar will not get to sit in his own seat.
  2. If any of the first 99 people sit in Bruna's seat, neymar will get to sit in his seat.

Probability Space: (Ω, F, P)

<u>Sample Space (Ω)</u>: Set of all possible outcomes.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

<u>Event</u> is a Subset of Sample Space

F is the set of all events (collection or set of sets)

F is the σ-algebra of Ω.

[F is a collection i.e. set of sets]

1.) $\Omega \in F$ ✓

2.) If $A \in F$ and $B \in F$, then $(A-B) \in F$

3.) If $A_i \in F$ for $i = 1, 2 \ldots n$, then $\bigcup_{i=1}^{n} A_i \in F$

$\Rightarrow$ $A \in F$ then $A^c \in F$

De Morgans Law:
Closed under intersection
$(\cup A)^c = (\cap A^c)$

Eg. $\Omega = \{1, 2, 3, 4, 5, 6\}$

One possible σ-algebra $\{\{1, 2, 3, 4, 5, 6\}, \{\phi\}\}$

Another possible σ-algebra $\{\Omega, \{\phi\}, \{1, 3, 5\}, \{2, 4, 6\}\}$

σ - algebra of Ω :

1.) $\Omega \in F$

2.) If $A \in F$ and $B \in F$ then $(A - B) \in F$

3.) If $A_i \in F$ then $\bigcup_i A_i \in F$

$\uparrow$ countable

$F = 2^{\Omega}$ (Power Set)

Set of all possible subsets of Ω

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$\{\Omega, \{\phi\}, \{1\}, \{2\} \ldots \{6\}$

$\{1, 2\} \ldots \{\ \}$

$\}$

# of elements in this will be 2^6

$\Omega = \underbrace{\{H, T\}}_{2}$; $2^{\Omega} = \{\{H, T\}, \{\phi\}, \{H\}, \{T\}\}$

# of elements are $2^2 = 4$

Probability is a function $P : F \rightarrow \mathbb{R}$

F ↑ Domain (Set)

$\mathbb{R}$ ↑ range Real Number

Three Axioms of probability :

1.) $P(A) \geq 0 \quad \forall A \in F$

2.) If $A_i \in F$ for $i = 1, 2, \ldots$

and $A_i \cap A_j = \phi \quad \forall i \neq j$ (mutually Exclusive)

then $P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + P(A_2) \ldots\ldots + P(A_n)$ (σ-additivity)

3.) $P(\Omega) = 1$

Eg: $\Omega = \{1, 2, \ldots 6\}$ Rolling of a dice

$P\{1\} = 1/6$, $P(\{1, 3, 5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{3}{6}$

Given the Axioms of Probability we can show:

1.) $0 \leq P(A) \leq 1 \quad \forall A \in \mathcal{F}$

$P(A) + P(A^c) = P(A \cup A^c)$
$= P(\Omega) = 1$

2.) $P(A^c) = 1 - P(A)$

3.) If $A \subset B$ (Subset) then, $P(A) \leq P(B)$

$\because B = A \cup (B \cap A^c) \Rightarrow P(B) = P(A) + \underbrace{P(B \cap A^c)}_{\geq 0}$

Mutually Exclusive

$\Rightarrow P(B) \geq P(A)$

Eg: Can Ω be countably infinity.

$\Omega = \{\omega_1, \omega_2 \cdots \infty\}$? $\Omega = \{1, 2, 3, \ldots\}$ set of all natural numbers

Let there be a number q such that $0 < q < 1$

$P(\{k\}) = q^{k-1}(1-q)$

$P(\Omega) = 1$? $P(\Omega) = \sum_{k=1}^{\infty} q^{k-1}(q-1)$

$= \frac{1}{q-1}(q-1) = 1$

Conditional Probability:

Given A and $B \in F$ and $P(B) > 0$

we define Cond. Prob of A given B as

$$P_B(A) = P(A|B) = \frac{P(AB)}{P(B)} \quad \Big\} \quad AB \Rightarrow A \cap B$$

mutually Exclusive

$$P_B(A_1 + A_2) = P_B(A_1) + P_B(A_2)$$

$$P_B(A_1 + A_2) = \frac{P((A_1 + A_2) \cap B)}{P(B)} = \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)}$$

Total Prob. formula: $P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)$

Total Prob. formula:

$$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)$$

$$\frac{P(AB)}{\cancel{P(B)}} \cdot \cancel{P(B)}^{\,1}, \qquad \frac{P(AB^c)}{\cancel{P(B^c)}} \cdot \cancel{P(B^c)}^{\,1},$$

$$P(AB) + P(AB^c) = P(AB + AB^c) = P(A)$$

Q.E.D.

Generalization:

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$ if $B_1, B_2 \ldots B_n$ are mutually exclusive and $B_1 \cup B_2 \cup \ldots B_n = \Omega$

Independent Events: Two events A and B are said to be independent

if $P(AB) = P(A) \cdot P(B)$

If $P(B) > 0$, then $P(A|B) = P(A)$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)\cancel{P(B)}}{\cancel{P(B)}} = P(A)$$

Eg: Rolling of two dice. Find prob. that both turn up six given the first one is six.

A → Both are six | A' → second one is six

B → First one is six

$\Omega = \{(1,1), (\ldots), \ldots (6,6)\}$; $A = \{(6,6)\}$; $B = \{(6,1), (6,2) \ldots (6,6)\}$

$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{1/36}{1/6} = \frac{1}{6} \Rightarrow P(A|B) \neq P(A) \Rightarrow$ A and B are not independent.

Bayes Rule:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B)}$$

$$= \frac{P(B|A)\;P(A)}{\underbrace{P(B|A)\,P(A) + P(B|A^c)\,P(A^c)}_{P(B)}}$$

Total Prob.

Bayes Rule:

eg: $C_1 \Rightarrow$ disease present

$C_2 \Rightarrow$ disease absent

$M \Rightarrow$ Test is +ve

$$P(C_1 | M) = \frac{P(M|C_1) \cdot P(C_1)}{\underbrace{P(M)}_{P(M|C_1)P(C_1) + P(M|C_2)P(C_2)}}$$

$$= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.1 \times (\underbrace{1 - 0.01}_{0.99})}$$

This is most sensitive term

Decide on C_1 if

$$P(C_1 | M) > P(C_2 | M)$$

or

$$\frac{P(C_1 | M)}{P(C_2 | M)} > \uparrow p$$

Cost of misclassification

Odds likelihood:

$$O(A) = \frac{P(A)}{1 - P(A)}$$

$$\frac{P(C_1|m)}{\underbrace{1 - P(C_1|m)}_{P(C_2|m)}} = \frac{P(m|C_1)}{P(m|C_2)} \cdot \frac{P(C_1)}{P(C_2)}$$

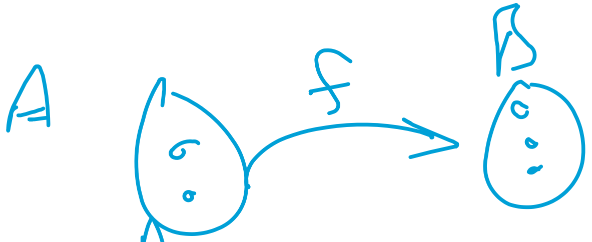
$$\underbrace{O(C_1|M)}_{\text{Posterior odds}} = \underbrace{\frac{P(m|C_1)}{P(M|C_2)}}_{\text{likelihood ratio}} \cdot \underbrace{O(C_1)}_{\text{Prior odds}}$$

Random Variable:

Given $(\Omega, \mathcal{F}, P)$, a random variable is a function mapping Ω to $\mathbb{R}$

rv, $X : \Omega \to \mathbb{R}$

s.t. $X^{-1}(B) \in \mathcal{F}$ $\forall B \in \mathcal{B}$ Borel-σalgebra



Given any function $f : A \to B$

set inverse is defined as $f : f^{-1} : 2^B \to 2^A$

f^{-1} maps subsets of B to subsets of A

$f^{-1}(C) = \{ a \in A : f(a) \in C \}$ $\quad C \subseteq B$

$X^{-1}(B) = \{ \omega \in \Omega : X(\omega) \in B \} \subseteq \Omega$

Transformation of $(\Omega, \mathcal{F}, P)$ to new prob. space

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (\mathbb{R}, \mathbb{B}, P_X)$$

$$\mathbb{B} \subseteq 2^{\mathbb{R}}$$

Eg: $\Omega = \{(i_1, i_2) : i_1, i_2 \in \{1, 2, 3, 4, 5, 6\}\}$

$X : \Omega \to \mathbb{R}$

$X(i_1, i_2) = i_1 + i_2$ [Can you have any other mapping?]

$P(X \leq 7) = P\{(1,1), (1,2) \dots\}$

$\mathbb{B} = (-\infty, 7] \Rightarrow X^{-1}(\mathbb{B}) = \{(1,1), (1,2), \dots\}$

$P_X(\mathbb{B}) = P[X^{-1}(\mathbb{B})] = P\{(1,1), (1,2) \dots\}$

Borel Subsets of $\mathbb{R}$

Given any set Ω and any $A \subseteq 2^{\Omega}$, we can define the smallest σ-algebra containing A as

$$\sigma(A) = \{ A, A^c, \Omega, \{\phi\} \}$$

Borel subsets of $\mathbb{R}$ is the smallest σ-algebra containing all sets of the form $(-\infty, x]$ for $x \in \mathbb{R}$.

$$\mathbb{B} = \sigma\left(\{ (-\infty, x] : x \in \mathbb{R} \right)$$

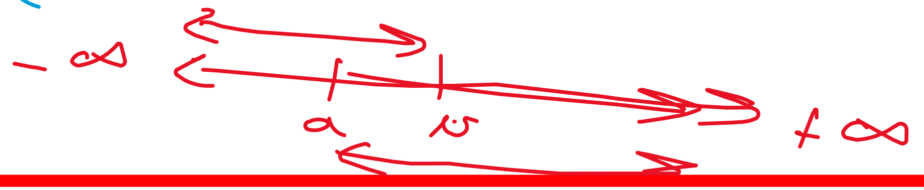
$$\mathbb{B} = \sigma\left(\{(-\infty, x] : x \in \mathbb{R}\}\right)$$

1.) $(-\infty, x] \in \mathbb{B} \quad \forall x \in \mathbb{R}$

2.) $(x, \infty) \in \mathbb{B} \quad \forall x \in \mathbb{R}$

3.) $(a, b] \in \mathbb{B} \quad \forall a, b \in \mathbb{R}$ and $a < b$

$$\bigcap \begin{cases} (-\infty, b] \in \mathbb{B} \\ (a, \infty) \in \mathbb{B} \end{cases}$$



4.) $[a, b] \in \mathbb{B}$

$(a - \frac{1}{n}, b] \in \mathbb{B}, \ \forall n$ and

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right]$$

5.) $[a, b) \in \mathbb{B}$

$[a, b - \frac{1}{n}] \in \mathbb{B} \ \forall n$ and $[a, b)$

$$= \bigcup_{n=1}^{\infty} \left[a, b - \frac{1}{n}\right]$$

6.) $(a, b) \in \mathbb{B}$

$$(\Omega, F, P) \xrightarrow{X} (\mathbb{R}, \mathbb{B}, P_X)$$

$$P_X(B) = P[X^{-1}(B)]$$

If X is a r.v. and $\mathbb{B}$ is a borel set

$$[x \in B] = \{\omega : X(\omega) \in B\}$$

A good way to represent/capture P_X is by using dist function

Cumulative distribution function of a r.v.: Given a r.v. X, the cdf of X, F_X is a function $F_X : \mathbb{R} \to \mathbb{R}$

$$F_X(x) = P_X((-\infty, x])$$
$$= P\{\omega : X(\omega) \in (-\infty, x]\}$$
$$= P\{\omega : X(\omega) \leq x\}$$

1.) $0 \leq F_X(x) \leq 1 \quad \forall x$

2.) $F_X(-\infty) = 0, \quad F_X(+\infty) = 1$

3.) F_X is monotone non-decreasing

4.) F_X is right continuous

$$\lim_{x_n \downarrow x} F_X(x_n) = F_X(x)$$

$x \Leftarrow x_n$

5.) F_X has a left hand limit

If $x_n \uparrow x$

$$\lim_{x_n \uparrow x} F_X(x_n) \text{ exists}$$

6.) $F_X(x) - \lim_{x_n \uparrow x} F_X(x_n)$

$$= P[X \leq x] - P[X < x]$$
$$= P\{X = x\}$$

$\{\omega : X(\omega) = x\}$

Discrete r.v : A r.v X is called a discrete r.v if there is a countable set $E \subseteq \mathbb{R}$ s.t.

$$P[X \in E] = 1$$

$$P\{\omega : X(\omega) \in E\} \text{ or } P_X(E)$$

If X is discrete then we denote E as

$$X \in \{x_1, x_2 \ldots \ldots\}$$

Prob. Mass Function : (pmf)

$$f_X(x) = P[X = x] \, . \; x \in E$$

$$= 0 \quad \forall x \notin E$$

Continuous r.v.: A r.v X is called continuous if df F_X is absolutely continuous or if there exists a function f_X s.t.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \quad \forall x \Bigg\} \Rightarrow f_X(x) = \frac{d F_X(x)}{dx}$$

$\forall x$ where f_X is continuous.

The f_X that exists is called the Prob. density function (pdf)

1.) $f_X(x) \geq 0 \quad \forall x$

2.) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

Random Variable: Expected Value

- Expected value : E(X)
  - Weighted average, of all possible values, considering their probabilities

- For Discrete random variable

$$\mathrm{E(X)} = \mu_X = \sum_{all\, values\, of\, x} [x.p(x)]$$

- For Continuous random variable

$$\mathrm{E(X)} = \mu_X = \int_{-\infty}^{\infty} x.f(x)dx$$

Random Variable: Variance and Standard Dev

- Variance of a Random Variable $\sigma_X^2 = E[(X-\mu_X)^2] = E(X^2) - \mu_X^2$

- Standard Deviation of a Random Variable $\sigma_x = +\sqrt{\sigma_X^2}$

- Where $E(X^2) = \sum_{all\,values\,of\,x} [x^2.p(x)]$, in the discrete case, and

$$E(X^2) = \int_{-\infty}^{\infty} x^2.f(x)dx \text{, in the continuous case}$$

- Variance and Standard Deviation reflects the extent to which the Random variable is close to its mean

The Mean

- For the Population

$$\mu = (x_1 + x_2 + \ldots + x_N) / N = \sum x_i / N$$

- For the Sample

$$\overline{x} = (x_1 + x_2 + \ldots. + x_n) / n \;=\; \sum x_i / n$$

The Variance and Standard Deviation (Population)

- Variance (Population)

$$\sigma^2 = \frac{\sum_{i=1}^{N} (Y_i - \mu)^2}{N}$$

- Standard Deviation (Population)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \mu)^2}{N}}$$

The Variance and Standard Deviation (Sample)

- Variance (Sample)

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

- Standard Deviation (Sample)

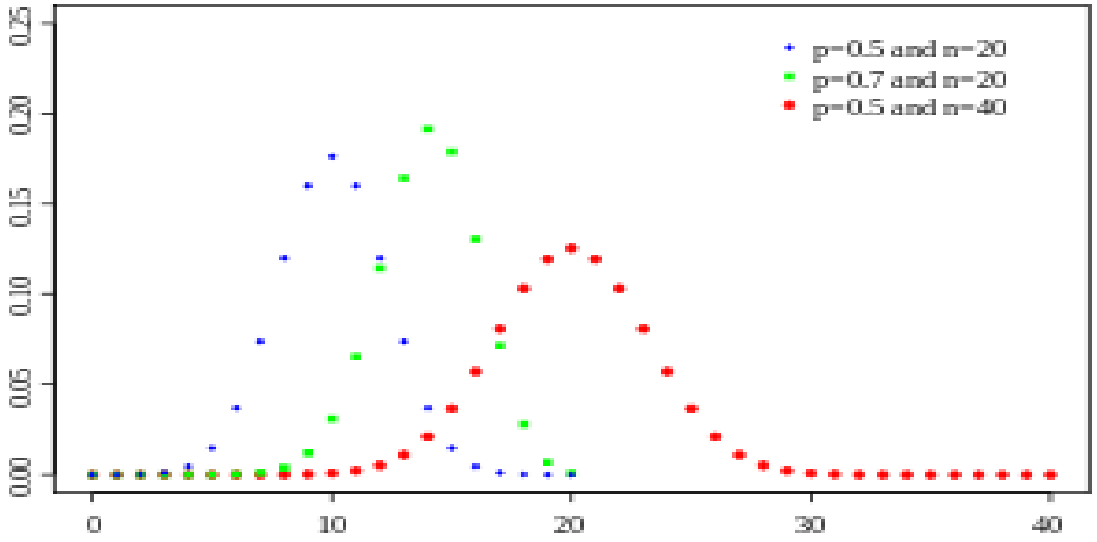
$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

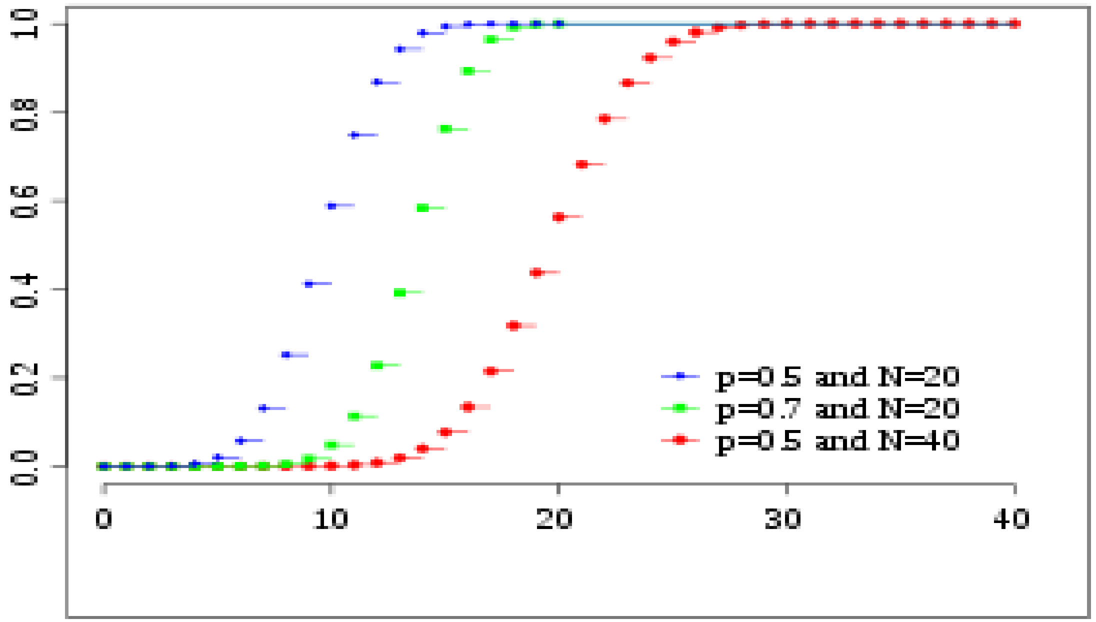
Binomial Distribution

- If an experiment has **only two outcomes**, it is known as a **Bernoulli Trial (or binomial trial)**
- Such an experiment is said to have Binomial Probability Distribution, if:
  - There are finite, **independent**, trial
  - Probability of success / failure is constant throughout the experiment
  - We are interested in the **number** of successes ‘x’, regardless of how they occur
- The number of successes is given by:

$$P(X = x) = P(x) = {}_nC_x\, p^x(1 - p)^{n - x} = \binom{n}{x} p^x(1 - p)^{n - x},\ x = 0, 1, 2, \ldots, n$$

Binomial Distribution





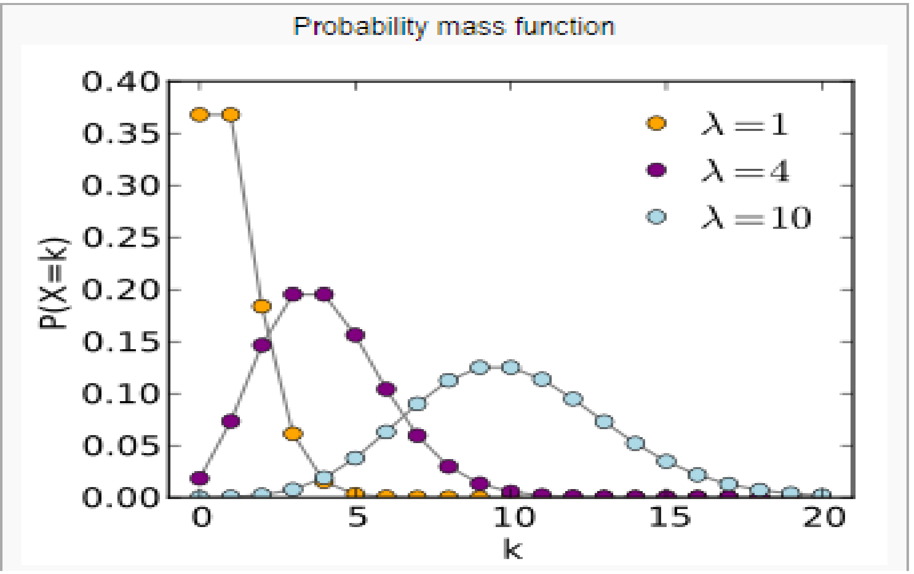
Notation	$B(n, p)$
Parameters	$n \in \mathbf{N}_0$ — number of trials $p \in [0,1]$ — success probability in each trial
Support	$k \in \{0, \ldots, n\}$ — number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lfloor (n+1)p \rfloor - 1$
Variance	$np(1-p)$
Skewness	$\dfrac{1-2p}{\sqrt{np(1-p)}}$

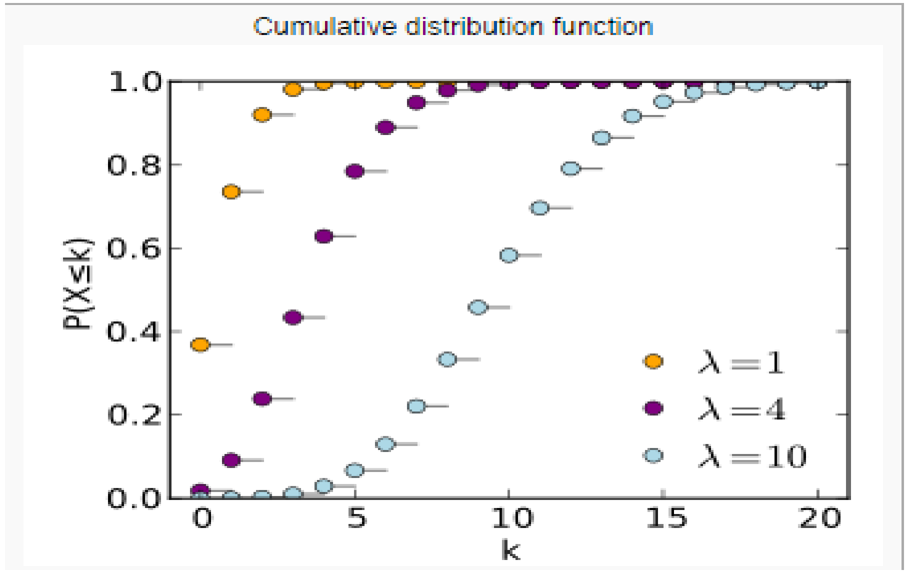
Src: Wikipedia

Poisson Distribution

- In case of random phenomena
- Where events are continuous
  - Calls arriving at a switchboard during lunch
  - Accidents at an intersection between 10 and noon
  - Misprints per page in a book
- The probability that a continuous measure X will take on value x, in a given unit of measurement, is governed by Poisson Distribution

Poisson Distribution





Notation	$\mathrm{Pois}(\lambda)$
Parameters	λ > 0 (real)
Support	$k \in \{0, 1, 2, 3, \ldots\}$
pmf	$\frac{\lambda^k}{k!}e^{-\lambda}$
CDF	$\frac{\Gamma(\lfloor k+1\rfloor,\lambda)}{\lfloor k\rfloor!}$, or $e^{-\lambda}\sum_{i=0}^{\lfloor k\rfloor}\frac{\lambda^i}{i!}$, or $Q(\lfloor k+1\rfloor,\lambda)$ (for $k \geq 0$, where $\Gamma(x,y)$ is the incomplete gamma function, $\lfloor k\rfloor$ is the floor function, and Q is the regularized gamma function)
Mean	λ
Median	$\approx \lfloor\lambda + 1/3 - 0.02/\lambda\rfloor$
Mode	$\lceil\lambda\rceil - 1, \lfloor\lambda\rfloor$
Variance	λ
Skewness	$\lambda^{-1/2}$

Src: Wikipedia

Geometric Distribution

- If in an experiment there are only two outcomes: Success | Failure
  - $P(s) = p$; $P(f) = q$;
  - $p + q = 1$
- We are interested in:
  - Number of trials 'x' to get the first success
- Geometric Distribution governs this case

$$P(X = x) = p\ .q^{x-1}, \qquad x = 1, 2, 3, \ldots$$

$$\mu = 1/p \text{ and } \sigma^2 = (1 - p)/p^2$$

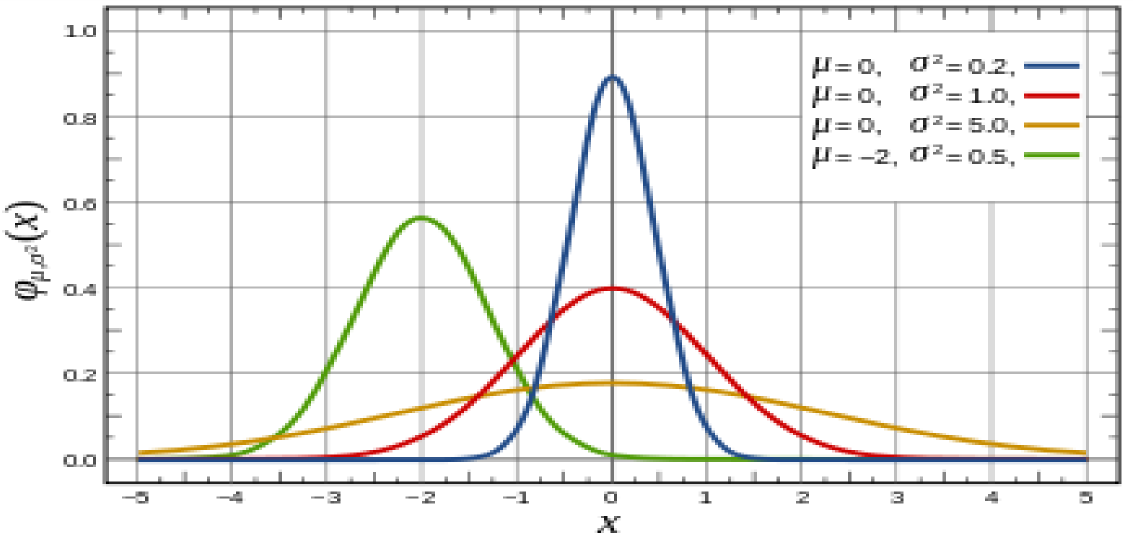
Normal Distribution

- A very well known Continuous Probability Distribution Function (PDF)
  - Applies to many phenomena
    - Human characteristics, physical quantities and processes, errors in physical and econometric measurements
  - Provides accurate approximation to a large number of probability laws
  - Important role in statistics and inferences
- PDF

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

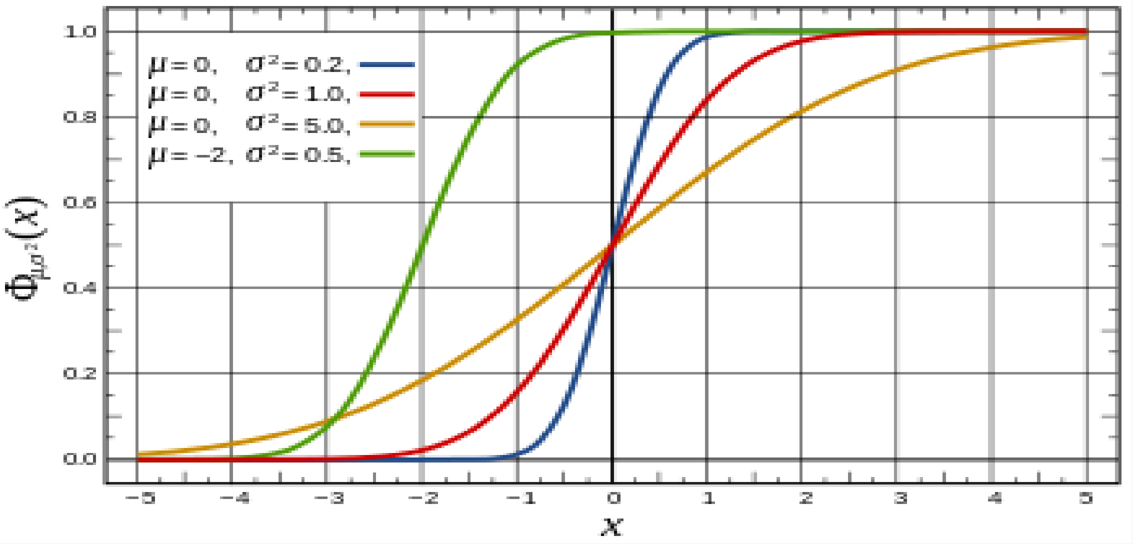
Normal Distribution

Probability density function



The red curve is the *standard normal distribution*

Cumulative distribution function



Notation	$\mathcal{N}(\mu,\,\sigma^2)$
Parameters	$\mu \in \mathbf{R}$ — mean (location) $\sigma^2 > 0$ — variance (squared scale)
Support	$x \in \mathbf{R}$
pdf	$\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu+\sigma\sqrt{2}\,\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
Skewness	0
Ex. kurtosis	0

Src: Wikipedia

Some common Probability Distributions

- Binomial Distribution
- Poisson Distribution
- Student's T-Distribution
- Chi-Square Distribution
- F-Distribution
- Normal Distribution
- Log normal Distribution
- Bernoulli Distribution
- Geometric Distribution
- Hypergeometric Distribution
- Multinomial Distribution
- Exponential Distribution
- Beta Distribution
- Gamma Distribution

Functions of random variables

Functions of random variables

A new random variable Y can be defined by applying a real Borel measurable function $g\colon \mathbb{R} \to \mathbb{R}$ to the outcomes of a real-valued random variable X. That is, $Y = g(X)$. The cumulative distribution function of Y is then

$$F_Y(y) = \mathrm{P}(g(X) \le y).$$

If function g is invertible (i.e., $h = g^{-1}$ exists, where h is g's inverse function) and is either increasing or decreasing, then the previous relation can be extended to obtain

$$F_Y(y) = \mathrm{P}(g(X) \le y) = \begin{cases} \mathrm{P}(X \le h(y)) = F_X(h(y)), & \text{if } h = g^{-1} \text{ increasing,} \\ \\ \mathrm{P}(X \ge h(y)) = 1 - F_X(h(y)), & \text{if } h = g^{-1} \text{ decreasing.} \end{cases}$$

With the same hypotheses of invertibility of g, assuming also differentiability, the relation between the probability density functions can be found by differentiating both sides of the above expression with respect to y, in order to obtain[5]

$$f_Y(y) = f_X\big(h(y)\big) \left| \frac{dh(y)}{dy} \right|.$$

Functions of random variables

Example Let X be a random variable with support $R_X = [1,2]$ and distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{2}x & \text{if } 1 \leq x \leq 2 \\ 1 & \text{if } x > 2 \end{cases}$$

Let

$$Y = X^2$$

The function $g(x) = x^2$ is strictly increasing and it admits an inverse on the support of X:

$$g^{-1}(y) = \sqrt{y}$$

The support of Y is $R_Y = [1,4]$. The distribution function of Y is

$$F_Y(y) = \begin{cases} 0 & \text{if } y < \chi, \forall \chi \in R_Y, \text{ i.e. if } y < 1 \\ F_X\left(g^{-1}(y)\right) = \frac{1}{2}\sqrt{y} & \text{if } y \in R_Y, \text{ i.e. if } 1 \leq y \leq 4 \\ 1 & \text{if } y > \chi, \forall \chi \in R_Y, \text{ i.e. if } y > 4 \end{cases}$$

Strictly increasing functions of a discrete random variable

When X is a discrete random variable, the probability mass function of $Y = g(X)$ can be computed as follows.

Proposition (probability mass of an increasing function) Let X be a discrete random variable with support R_X and probability mass function $p_X(x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be strictly increasing on the support of X. Then, the support of $Y = g(X)$ is

$$R_Y = \{y = g(x) : x \in R_X\}$$

and its probability mass function is

$$p_Y(y) = \begin{cases} p_X\left(g^{-1}(y)\right) & \text{if } y \in R_Y \\ 0 & \text{if } y \notin R_Y \end{cases}$$

Strictly increasing functions of a discrete random variable

Example Let X be a discrete random variable with support

$$R_X = \{1, 2, 3\}$$

and probability mass function

$$p_X(x) = \begin{cases} \frac{1}{6}x & \text{if } x \in R_X \\ 0 & \text{if } x \notin R_X \end{cases}$$

Let

$$Y = g(X) = 3 + X^2$$

The support of Y is

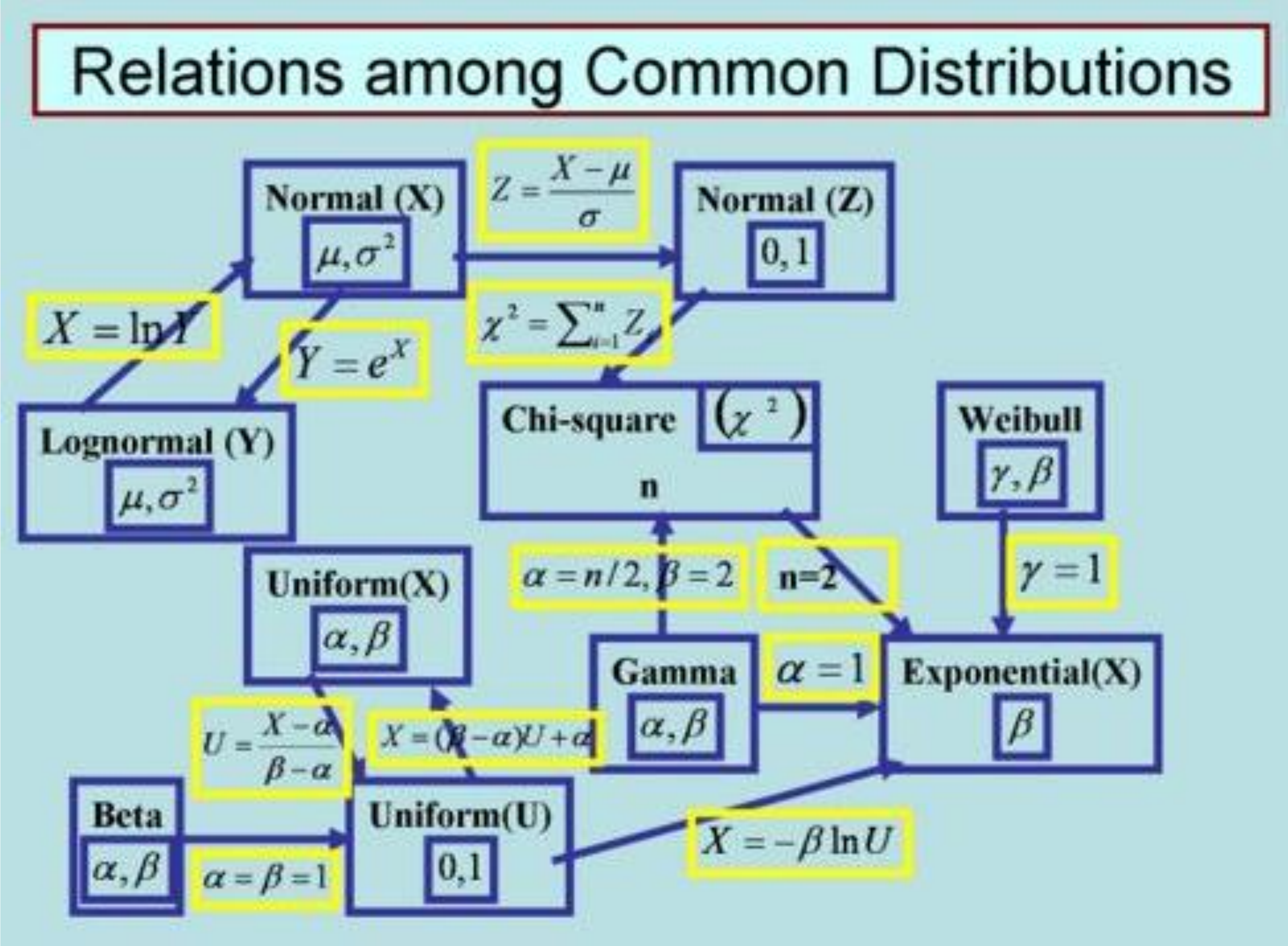
$$R_Y = \{4, 7, 12\}$$

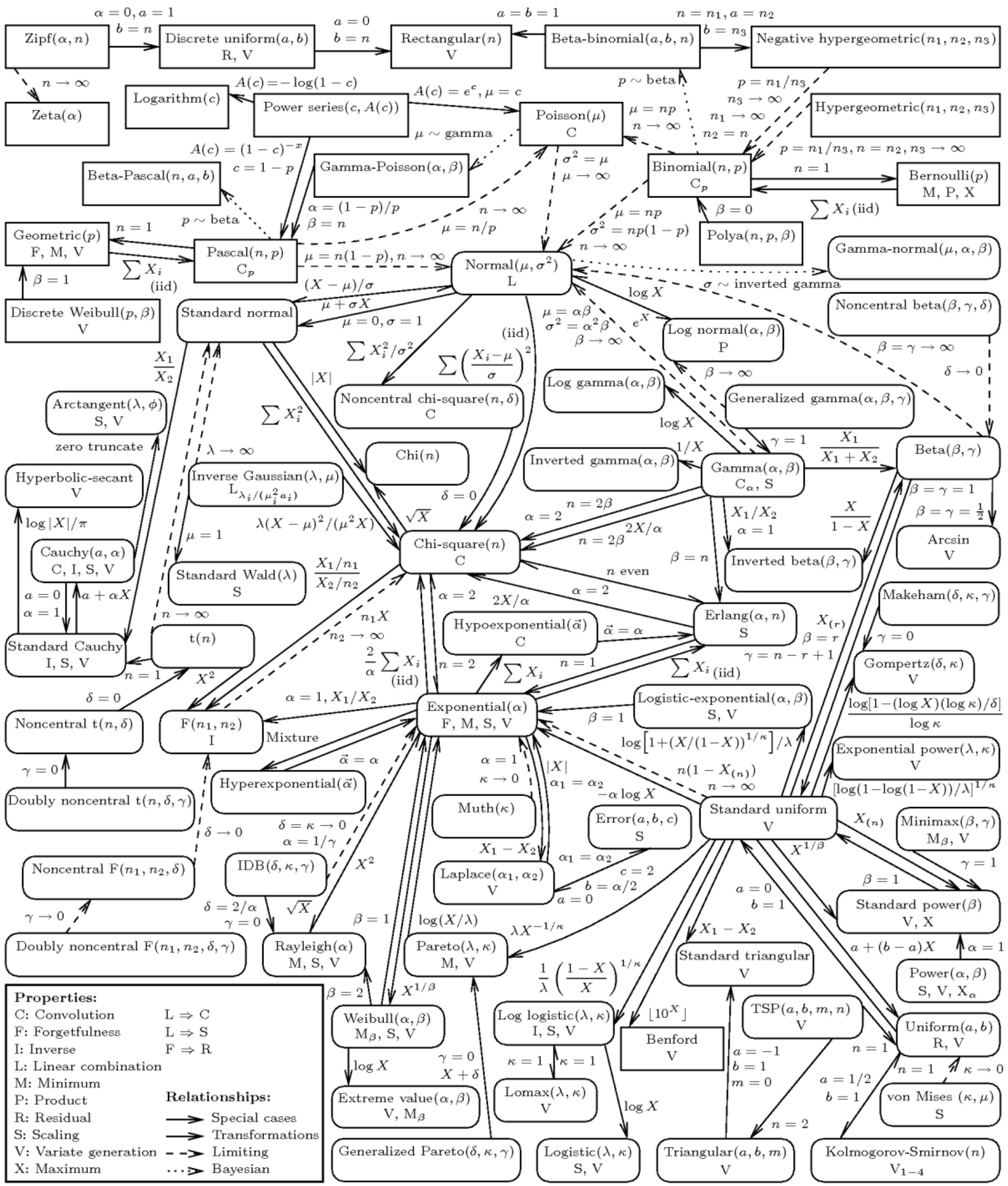
The function g is strictly increasing and its inverse is

$$g^{-1}(y) = \sqrt{y - 3}$$

The probability mass function of Y is

$$p_Y(y) = \begin{cases} \frac{1}{6}\sqrt{y-3} & \text{if } y \in R_Y \\ 0 & \text{if } y \notin R_Y \end{cases}$$


Relations among Common Distributions
Normal (X)
μ, σ^2
$Z = \frac{X-\mu}{\sigma}$
Normal (Z)
0,1
$X = \ln Y$
$Y = e^X$
$\chi^2 = \sum_{i=1}^{n} Z_i$
Lognormal (Y)
μ, σ^2
Chi-square (χ^2)
n
Weibull
γ, β
$\alpha = n/2, \beta = 2$
n=2
$\gamma = 1$
Uniform(X)
α, β
Gamma
α, β
$\alpha = 1$
Exponential(X)
β
$U = \frac{X-\alpha}{\beta-\alpha}$
$X = (\beta-\alpha)U + \alpha$
Beta
α, β
$\alpha = \beta = 1$
Uniform(U)
0,1
$X = -\beta \ln U$


Zipf(α, n)
Discrete uniform(a, b) R, V
Rectangular(n) V
Beta-binomial(a, b, n)
Negative hypergeometric(n1, n2, n3)
Zeta(α)
Logarithm(c)
Power series(c, A(c))
Poisson(μ) C
Hypergeometric(n1, n2, n3)
Beta-Pascal(n, a, b)
Gamma-Poisson(α, β)
Binomial(n, p) Cp
Bernoulli(p) M, P, X
Geometric(p) F, M, V
Pascal(n, p) Cp
Normal(μ, σ²) L
Polya(n, p, β)
Gamma-normal(μ, α, β)
Discrete Weibull(p, β) V
Standard normal
Log normal(α, β) P
Noncentral beta(β, γ, δ)
Noncentral chi-square(n, δ) C
Log gamma(α, β)
Generalized gamma(α, β, γ)
Arctangent(λ, φ) S, V
Chi(n)
Inverted gamma(α, β)
Gamma(α, β) Cα, S
Beta(β, γ)
Hyperbolic-secant V
Inverse Gaussian(λ, μ)
Chi-square(n) C
Arcsin V
Cauchy(a, α) C, I, S, V
Standard Wald(λ) S
Inverted beta(β, γ)
Makeham(δ, κ, γ)
Hypoexponential(α) C
Erlang(α, n) S
Standard Cauchy I, S, V
t(n)
Gompertz(δ, κ) V
Noncentral t(n, δ)
F(n1, n2) I
Exponential(α) F, M, S, V
Logistic-exponential(α, β) S, V
Exponential power(λ, κ) V
Doubly noncentral t(n, δ, γ)
Hyperexponential(α)
Muth(κ)
Error(a, b, c) S
Standard uniform V
Minimax(β, γ) Mβ, V
Noncentral F(n1, n2, δ)
IDB(δ, κ, γ)
Laplace(α1, α2) V
Standard power(β) V, X
Doubly noncentral F(n1, n2, δ, γ)
Rayleigh(α) M, S, V
Pareto(λ, κ) M, V
Standard triangular V
Power(α, β) S, V, Xα
Weibull(α, β) Mβ, S, V
Log logistic(λ, κ) I, S, V
Benford V
TSP(a, b, m, n) V
Uniform(a, b) R, V
Extreme value(α, β) V, Mβ
Lomax(λ, κ) V
von Mises (κ, μ) S
Generalized Pareto(δ, κ, γ)
Logistic(λ, κ) S, V
Triangular(a, b, m) V
Kolmogorov-Smirnov(n) V1−4
Properties:
C: Convolution
F: Forgetfulness
I: Inverse
L: Linear combination
M: Minimum
P: Product
R: Residual
S: Scaling
V: Variate generation
X: Maximum
L ⇒ C
L ⇒ S
F ⇒ R
Relationships:
Special cases
Transformations
Limiting
Bayesian