TYPES OF VARIABLE

NOTATION: p_X = PROB. MASS FUNCTION

P_X = PROBABILITY

- DISCRETE RANDOM VARIABLE
  - BERNOULLI
  - BINOMIAL
  - HYPERGEOMETRIC
  - POISSON
  - GEOMETRIC
- CONTINUOUS RANDOM VARIABLE
  - UNIFORM
  - BETA
  - EXPONENTIAL
  - GAMMA
  - GAUSSIAN (NORMAL)

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

BERNOULLI (DISCRETE) $X \sim Bern(p)$

- SINGLE TRIAL WITH 2 POSSIBLE OUTCOMES $\begin{cases} p & \text{SUCCESS} \\ 1-p & \text{FAILURE} \end{cases}$

EXAMPLE: $p = 0.4$ THEN:
- $P(1) = 0.4$
- $P(0) = 1 - 0.4 = 0.6$
- $E[X] = p = 0.4$

- PROBABILITY MASS FUNCTION: $f(x) = p^x (1-p)^{1-x} \quad x = 0,1$
- PROBABILITY: $P(1) = p; \quad P(0) = 1-p$
- EXPECTED VALUE: $E[X] = \underset{\text{VALUE}}{0} \cdot \underset{\text{PROB.}}{P(0)} + \underset{\text{VALUE}}{1} \underset{\text{PROB.}}{P(1)} = P(1) = p$

$$E[X^2] = 0^2 \cdot P(0) + 1^2 P(1) = P(1) = p$$

- VARIANCE: $VAR(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$

BINOMIAL (DISCRETE) $X \sim Bin(n,p)$

Used to indicate that the random variable X has the binomial distribution for positive integer parameter *n* and real parameter *p* satisfying *0 < p < 1*. It models the number of successes in *n* mutually independent Bernoulli trials, each with probability of success *p*.

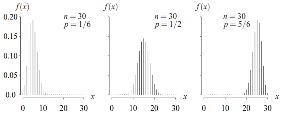
Can be use to model the number of people in a group of *n* people with a particular characterstic, the number of defective items in a batch of *n* items, the number of fours in *n* rolls of a fair die. Is <u>number of successes in *n* mutually independent Bernoulli trials</u>.

EXAMPLE: EXAM OF 4 QUESTIONS WITH 4 ANSWER EACH. PROBABILITY THAT EXACTLY 2 ARE CORRECT

$$X_i = \begin{cases} 1 & \text{IF ANSWER TO } i\text{-TH QUESTION IS CORRECT} \\ 0 & \text{IF ANSWER TO } i\text{-TH QUESTION IS WRONG} \end{cases}$$

$$P[S_4 = 2] = \binom{4}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^2 = \frac{54}{256}$$

$$p_{S_n}(x) = \binom{n}{k} p^k (1-p)^{n-k} \quad k \in \{0,1,\dots n\}$$



- PROBABILITY MASS FUNCTION: $f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x \in 0,1,\dots,n$
- EXPECTED VALUE:

$$E[S_n] = np$$

$$M(t) = E[e^{tX}] = (1-p+pe^t)^n \quad -\infty < t < +\infty$$

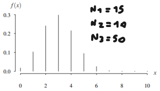
- VARIANCE: $VAR(X) = np(1-p)$

HYPERGEOMETRIC (DISCRETE) $X \sim \text{HYPERGEOM}(N_1, N_2, N_3)$

$N_1, N_2 \in \{0, 1, \dots N_3\}$

Is used for sampling without replacement from a finite population of items. A hypergeometric random variable X is the number of defective items in a sample of size n2 items drawn at random and without replacement from a lot of n3 items which contains n1 defective items.

Application for acceptance sampling from quality control and animal population size estimation using tagging with capture/recapture.

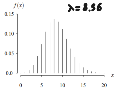


- **PROBABILITY MASS FUNCTION:** $f(x) = \dfrac{\binom{N_1}{x}\binom{N_3 - N_1}{N_2 - x}}{\binom{N_3}{N_2}}$
- **EXPECTED VALUE:** $E[x] = \dfrac{N_1 N_2}{N_3}$
- **VARIANCE:** $VAR(X) = \dfrac{N_2 N_1 (N_3 - N_1)(N_3 - N_2)}{N_3^2 (N_3 - 1)}$

POISSON (DISCRETE) $X \sim \text{POISSON}(\lambda)$

Can be used to model the number of events in an interval associated with a process that evolves randomly over space or time.

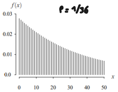
Applications include the number of potholes (buche) over a stretch of highway, the number of typographical errors in a book, the number of customer arrivals in an hour, the number of earthquakes in a decade. Can also be used to approximate the binomial distribution when *n* is large and *p* is small



- **PROBABILITY MASS FUNCTION:** $f(x) = e^{-\lambda}\dfrac{\lambda^k}{k!}$
- **EXPECTED VALUE:** $E[x] = \lambda$
- **VARIANCE:** $VAR(x) = \lambda$

GEOMETRIC (DISCRETE) $X \sim \text{GEOM}(P)$ $0 < P < 1$

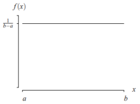
Can be used to model the number of failures before the first success in repeated mutually independent Bernoulli trials, each with probability of success *p*. The geometric distribution with *p* = 1/36 would be an appropriate model for the number of rolls of a pair of fair dice prior to rolling the first double six. Is the only discrete distribution with memoryless property (for continuous it is the exponential).



- **PROBABILITY MASS FUNCTION:** $f(x) = P(1-P)^x$ $\quad x = 0, 1, 2 \dots$
- **EXPECTED VALUE:** $E[x] = \dfrac{1-P}{P}$
- **VARIANCE:** $VAR(x) = \dfrac{1-P}{P^2}$

UNIFORM (CONTINOUS) $X \sim UNIF(a, b)$

Indicates that a random variable X has the uniform distribution with minimum *a* and maximum *b*. It is used to model a random variable that is equally likely to occur between *a* and *b*. It is an appropriate model for the position of the puncture on a flat tire. Event times associated with a Poisson process are uniformly distributed over an interval



- **PROBABILITY MASS FUNCTION:** $f(x) = \frac{1}{b-a}$

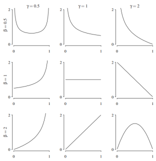
- **EXPECTED VALUE:** $E[x] = \frac{a+b}{2}$

- **VARIANCE:** $VAR(X) = \frac{(b-a)^2}{12}$

BETA (CONTINOUS) $X \sim BETA(\alpha, \beta)$

A beta random variable X with positive shape parameter α and β density function.

Used for modeling random variables that lie between 0 and 1 (percentages or interest rates) and as a prior distribution.



- **PROBABILITY MASS FUNCTION:** $\frac{\Gamma(\alpha+\beta)\, x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}$ $\quad 0 < x < 1$

- **EXPECTED VALUE:** $E[x] = \frac{\alpha}{\alpha+\beta}$

  $E[x^2] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$

- **VARIANCE:** $VAR(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

*** GAMMA**

- $\Gamma(z) = \int_0^{+\infty} e^{-t}\, t^{z-1}\, dt \quad z > 0$

- $\Gamma(1) = 1$
- $\Gamma(2) = 1!$
- $\Gamma(6) = 5!$

$\Gamma(n+1) = n!$

*** EULER'S THEOREM**

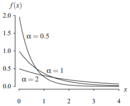
- $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

- $\int_0^1 x^6(1-x)^4\, dx = \frac{\Gamma(7)\Gamma(5)}{\Gamma(12)} = \frac{6!\,4!}{11!}$

EXPONENTIAL (CONTINOUS) $X \sim EXPONENTIAL(\alpha)$

The random variable X has the exponential distribution with positive scale parameter α. The exponential distribution can be parameterized by its *mean* α. The exponential distribution is used in reliability to model the lifetime of an object which (in statistical sense) does not age. Is the only continuous variable with memoryless property (the discrete one is the geometric).

Is used in queueing theory to model the times between customer arrivals and the service time. Is used to model the lifetime of an organism or the survival time after treatment.



- **PROBABILITY MASS FUNCTION:** $f(x) = \lambda e^{-\lambda x} \quad \lambda > 0$

- **EXPECTED VALUE:** $E[x] = 1/\lambda$

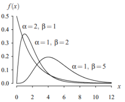
  $E[x^2] = \frac{2}{\lambda^2}$

- **VARIANCE:** $VAR(X) = \frac{1}{\lambda^2}$

GAMMA (CONTINOUS) $X \sim GAMMA(\lambda, \tau)$

A gamma random variable X with positive shape parameter λ and τ probability density function.

Can be used to model service times, lifetimes of objects and repair times. Has an exponential right-hand tail



- **PROBABILITY MASS FUNCTION:** $\frac{\lambda^{\tau}}{\Gamma(\tau)} x^{\tau-1} e^{-\lambda x} \quad x > 0$

- **EXPECTED VALUE:** $E[x] = \frac{\tau}{\lambda}$

  $E[x^2] = \frac{\tau(\tau+1)}{\lambda^2}$

- **VARIANCE:** $VAR(X) = \frac{\tau}{\lambda^2}$

REMEMBER IF $\lambda = 1/2$ AND $\tau = n/2$

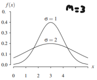
$$f(x) = \frac{1}{2^{n/2}\,\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

CHI-SQUARED FUNCTION WITH N DEGREES OF FREEDOM

GAUSSIAN (CONTINOUS) $X \sim N(m, \sigma^2)$ $\quad -\infty < m < +\infty \quad \sigma > 0$

A gaussian/normal random variable X with mean *m* and variance σ^2

Can be used for modeling adult heights, newborn baby weights, ball bearing diameters...The normal distribution can be used to approximate the binomial distribution when *n* is large and *p* is close to 1/2. The normal distribution can also be used to approximate the Poisson distribution when *n* is large and *p* is small.



m=3

- m: WHERE THE PEAK IS CENTERED
- SMALL σ^2: TALL AND SKINNY
- BIG σ^2: FLATTERING THE CURVE

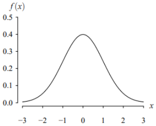
- **PROBABILITY MASS FUNCTION:** $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-m)^2}{2\sigma^2}}$

- **EXPECTED VALUE:** $E[x] = m$

- **VARIANCE:** $VAR(X) = \sigma^2$

STANDARD NORMAL (GAUSSIAN) (CONTINOUS) $X \sim N(0,1)$

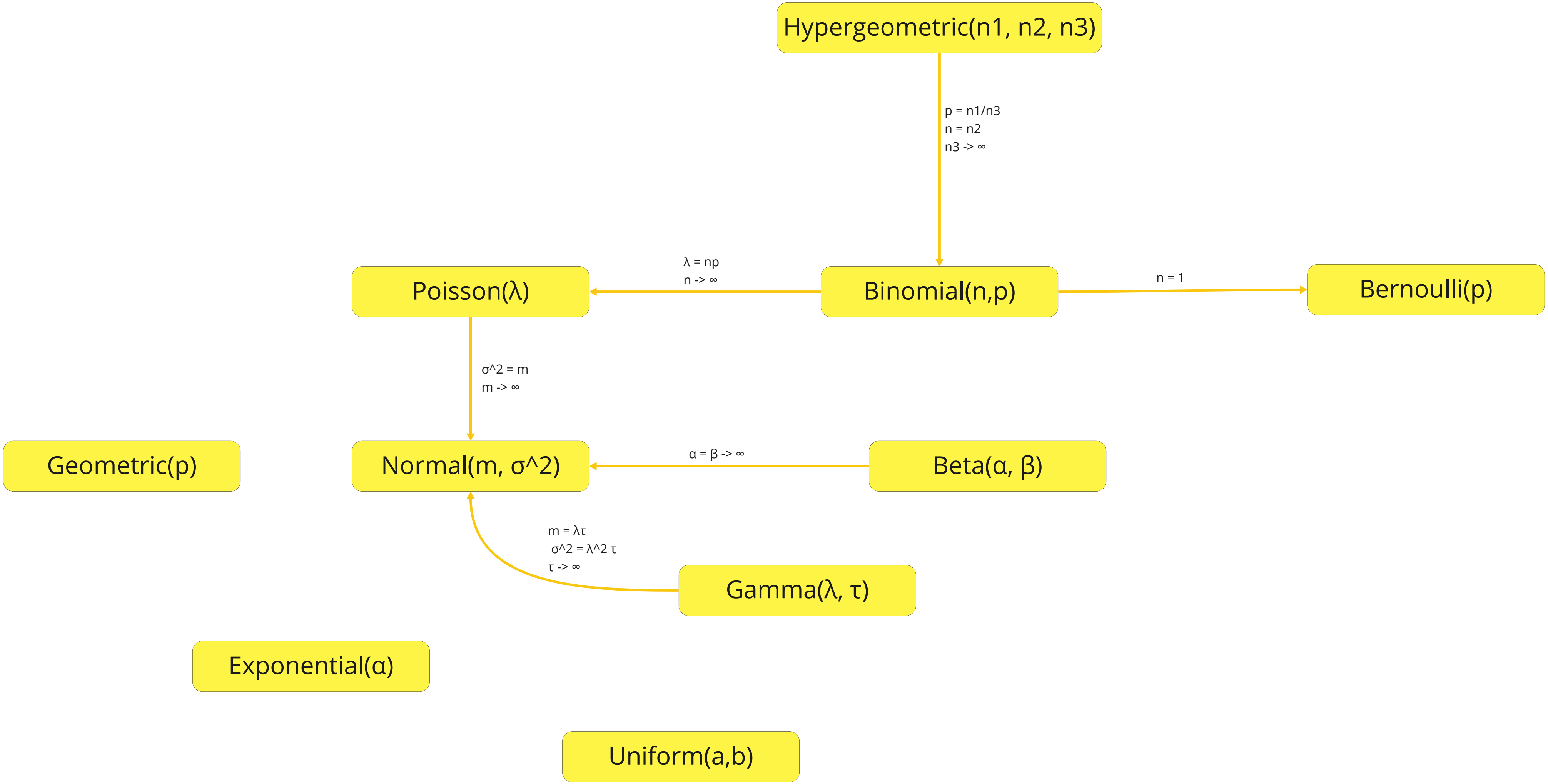
Obtained by putting m = 0 and σ^2 = 1. Only a single table is required for all calculations involving the normal distribution.



- **PROBABILITY MASS FUNCTION:** $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad x \in \mathbb{R}$

- **EXPECTED VALUE:** $E[x] = 0$

- **VARIANCE:** $VAR(X) = 1$



Hypergeometric(n1, n2, n3)
p = n1/n3
n = n2
n3 -> ∞
λ = np
n -> ∞
Poisson(λ)
Binomial(n,p)
n = 1
Bernoulli(p)
σ^2 = m
m -> ∞
Geometric(p)
Normal(m, σ^2)
α = β -> ∞
Beta(α, β)
m = λτ
σ^2 = λ^2 τ
τ -> ∞
Gamma(λ, τ)
Exponential(α)
Uniform(a,b)