4- random variables and distributions

discrete distributions

- $\boxed{\text{binomial / bernoulli}}$ = number of successes(k) in n bernoulli trials with p prob of success

↳ bernoulli random variable = outcome of trial 0 or 1 with probability p (failure ← 0, 1 → success)

robability of side effects is 10%, among 40 people what is the probability that exactly 3 people will have side effects? $\binom{40}{3} \cdot (0.1)^3 \cdot (0.9)^{37}$ $f(0.1, 3, 40)$

at most 1 person? $\binom{40}{0}(0.1)^0 \cdot (0.9)^{40} + \binom{40}{1}(0.1)^1 \cdot (0.9)^{39}$ $F(0.1, 1, 40)$

mass function = $f(p,k,n) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$

cummulative density function = $F(p,k,n) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$

mean = $\mu = np$ variance = $\sigma = np(1-p)$

- $\boxed{\text{geometric}}$ = number of trials (k) needed to get first success with p success

example = success probability is 0.2, probability of success at 3rd trial?

↳ $(0.8)^2 \cdot (0.2)$ $f(0.2, 2)$

↳ success at at most 3rd trial $P(x \leq 3) = 1-(0.8)^3$ $F(0.2, 2)$

mass function = $f(p,k) = (1-p)^k \cdot p$ $P(X = k+1)$

cummulative density function = $F(p,k) = 1-(1-p)^k$ $P(X \leq k)$

mean = $\frac{1-p}{p}$ variance = $\frac{1-p}{p^2}$

example = a cell divides once per hour, %10 probability of mutation at every mitosis. How many hours before there is at least %50 of mutation?

↳ 1 hour → $1-(0.9)^1 = 0.1 \rightarrow 1$ div $F(0.1, 2^1-1)$

2 hour → $1-(0.9)^3 = 0.27 \rightarrow 3$ div $F(0.1, 2^2-1)$

3 hour → $1-(0.9)^7 = 0.52 \rightarrow 7$ div total $F(0.1, 2^3-1)$

↳ $F(0.1, 2^k-1) \geq 0.5 = ?$

3 hours at least

- $\boxed{\text{negative binomial}}$ = probability of getting a certain number of successes (k) before a certain number of failures (r) with p success change → generalization of geometric dist.

example = success probability is 0.7, what is the probability of third success in fifth trial? $\binom{4}{2} \cdot (0.7)^2 \cdot (0.3)^2 \cdot 0.7 = \binom{4}{2} \cdot (0.3)^2 \cdot (0.7)^3$ ✓ ✗ ✓ ✗ ✓

mass func = $f(p,k,r) = \binom{k+r-1}{k} (1-p)^r p^k$ average = $rp / 1-p$

- $\boxed{\text{poisson}}$ = the probability of the number of events (k) recorded on a given timeframe T, if the average rate of the events are λ/k

example = avg 5% event occur every hour, out of 100, probability of 2 occurs?

↳ $\lambda = 5 \rightarrow f(5, 2)$

mass func = $f(\lambda, k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$ mean = variance = λ

continuous distributions

- $\boxed{\text{normal / gaussian}}$ = represents some random variable $x = \sum_i x_i$ of a large number of random variables. standard normal distribution is $\mu = 0$ and, std = var = 1

↳ z score = $\frac{x-\mu}{\sigma}$ shifted and scaled score to allow using the standard distribution

density function = $f(\mu, \sigma)$ cummulative density func = $F(\mu, \sigma)$

- $\boxed{\text{exponential}}$ = represents the duration (x) between two consecutive events in a poisson process with λ rate. → continuous version of geometric distribution

density func = $f(\lambda, x) = \lambda e^{-\lambda x}$ mean = $\mu = \lambda^{-1}$

cummulative density func = $F(\lambda, x) = 1 - e^{-\lambda x}$ variance = $\sigma^2 = \lambda^{-2}$

example = an event happen once a year ($\lambda = 1$), what is the probability that it will not happen within one year given that it happened today?

↳ $P(t>1) = 1 - F(1,1) = 0.368$ → same

example = what is the same probability given that it did not happen in a year?

↳ $P(t>2 \mid t>1) = P(t>2, t>1) / P(t>1) = P(t>2) / P(t>1) = (1-F(1,2)) / (1-F(1,1)) = 0.368$

memoryless = the probability of time-to-arrival t of an event is independent of how much time has passed since the last event $P(X > s+t \mid X > s) = P(X > t)$

↳ $P(X > s+t \mid X > s) \neq P(X > s+t)$

↳ only the exponential and geometric distributions are memoryles.

- $\boxed{\text{gamma}}$ = generalization of the exponential distribution with two parameters θ and k

↳ exponential dist is gamma dist when k = 1 and $\theta = 1/k$

density func = $f(x, \theta, k)$ cumulative density func = $F(x, \theta, k)$ mean = $k \cdot \theta$

example = event rate is 1 in every 2 minutes on average with exponentially distributed there are 5 events on the line waiting, probability of it takes at most 10 min?

↳ $\lambda = \frac{1}{2} = 0.5$ $\left(\frac{1}{\lambda} \ \frac{2\,\text{min}}{1\,\text{min}}\right)$ $k = 5$ $\theta = 2 \longrightarrow F(10, 2, 5)$

central limit theorem (CLT)

states that the distribution of a sample variable approximates a normal distribution as the sample size becomes larger, assuming random and independent sampling