Lab 2

Probability review

Pairwise independence does not imply mutual independence.

- X = Flip coin, Pr(X = heads) = ½
- Y = Flip Coin, Pr(Y = heads) = ½
- Z = Pr(exactly X or Y are heads (not both)) = ½

X	Y	Z = XOR(X,Y)
0	0	0
1	0	1
0	1	1
1	1	0

X	Y	Pr[X,Y]
0	0	1/4
1	0	1/4
0	1	1/4
1	1	1/4

X	Z	Pr[X,Y]
0	0	1/4
1	1	1/4
0	1	1/4
1	0	1/4

Pairwise independence does not imply mutual independence.

- Pr(X,Z) pairwise independence:
  - Pr(X=0,Z=0) = Pr(X=0)* Pr(Z=0) = ½ * ½ = ¼
- NOT mutual independence
  - Pr(X=1,Y=1,Z=0) = ¼ where as Pr(X=1)*Pr(Y=1)*Pr(Z=0) = 1/8

X	Y	Z = XOR(X,Y)	Pr[X]*Pr[Y]*Pr[Z]	Pr[X,Y,Z]
0	0	0	⅛	¼
1	0	1	⅛	¼
0	1	1	⅛	¼
1	1	0	⅛	¼
0	0	1	⅛	0
1	0	0	⅛	0
0	1	0	⅛	0
1	1	1	⅛	0

Definitions

- Expectation
  - When X is a discrete random variable
    - $\sum_i i \cdot \Pr[X = i]$
  - When X is a continuous random variable, and f(x) is its probability density function
    - $\int_{-\infty}^{\infty} x \cdot f(x) dx$
- Variance
  - $Var[X] = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2$
  - It measures how far your data spread out from their average value.
- Standard deviation
  - $\sigma_X = \sqrt{Var[X]}$

Covariance and Correlation

Definitions: $$Cov(X,Y) = E[(X - E[x])(Y - E[Y])]$$

$$= \sum_x \sum_y (x - E[X])(y - E[Y]) p_{X,Y}(x,y)$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

From a linear algebra perspective: If X, Y are two random variables of zero mean, then the covariance Cov[XY] = E[X · Y] is the dot product of X and Y . The standard deviation of X is the length of X. The correlation is the cosine of the angle between the two vectors.

Covariance

Some properties:

$$Cov(X + c, Y) = Cov(X, Y)$$

$$Cov(aX + bY, Z) = a \cdot Cov(X, Z) + b \cdot Cov(Y, Z)$$

Exercise:
X and Y are two independent $\mathcal{N}(0, 1)$ random variables and:

$$Z = 1 + X + XY^2$$

$$W = 1 + X$$

Find $Cov(W, Z)$

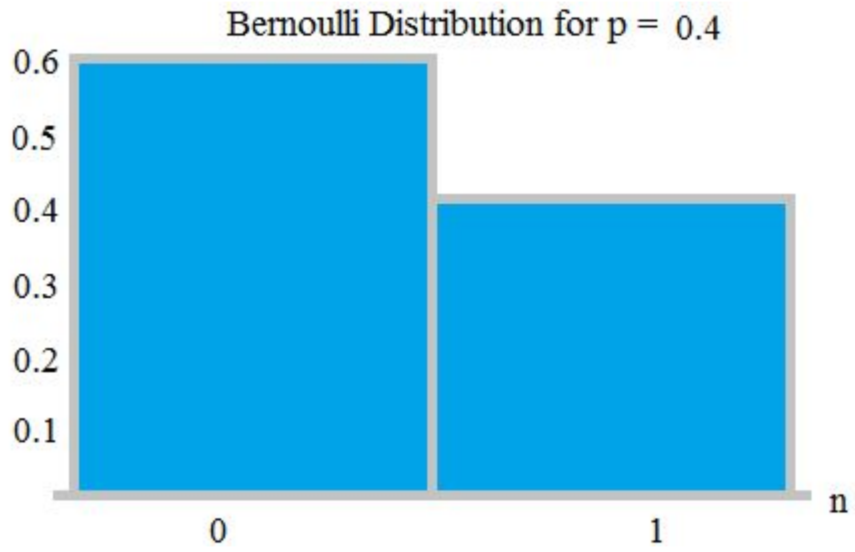
Discrete probability distributions

	PMF	Mean	Variance
Bernoulli (p)	1-p if k=0 p if k=1	p	p(1-p)
Binomial (n,p)	$\binom{n}{k} p^k (1-p)^{n-k}$	np	np(1-p)
Poisson (λ)	$\frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ
Geometric (p)	$(1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Bernoulli(p)

- Discrete
- Two outcomes: 0,1 || fail, success
- Pr(X = 1) = p = 1-q
- Pr(X = 0) = q = 1-p

- Expectation:
  - E[x] = 1*p + (0)*(1-p) = p



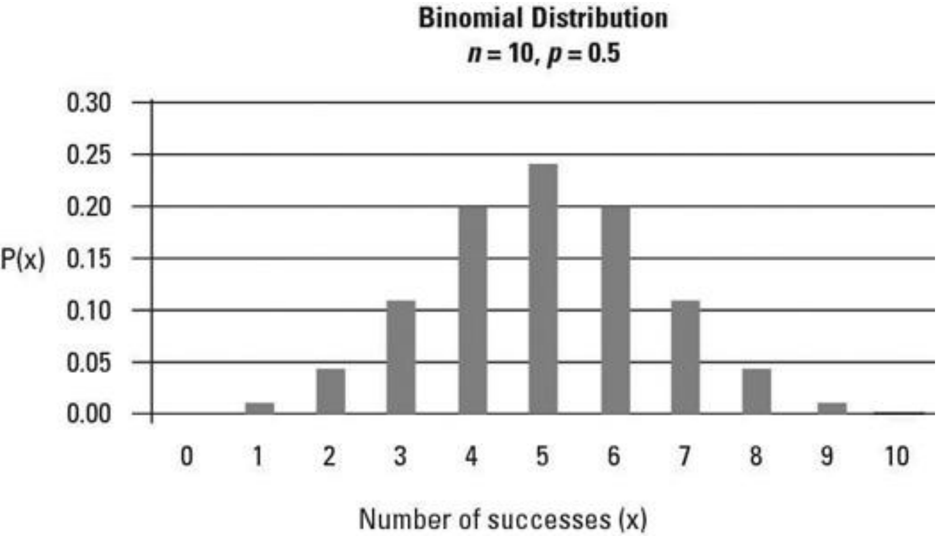
Binomial(n,p)

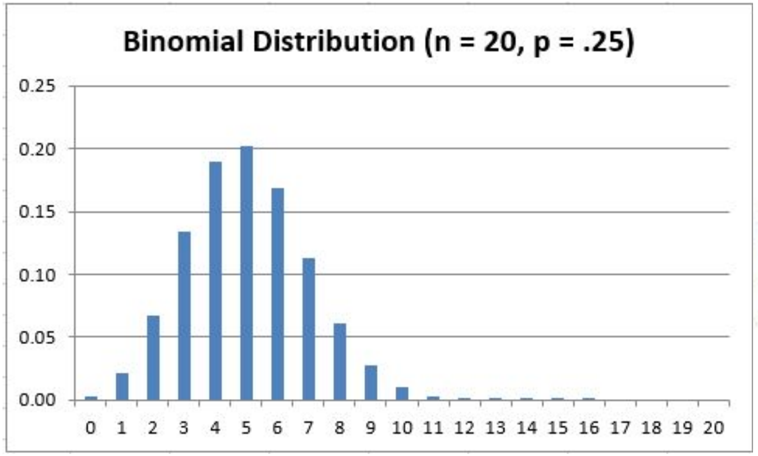
Example: Flip a coin 5 times, what is the probability of having exactly 2 heads?

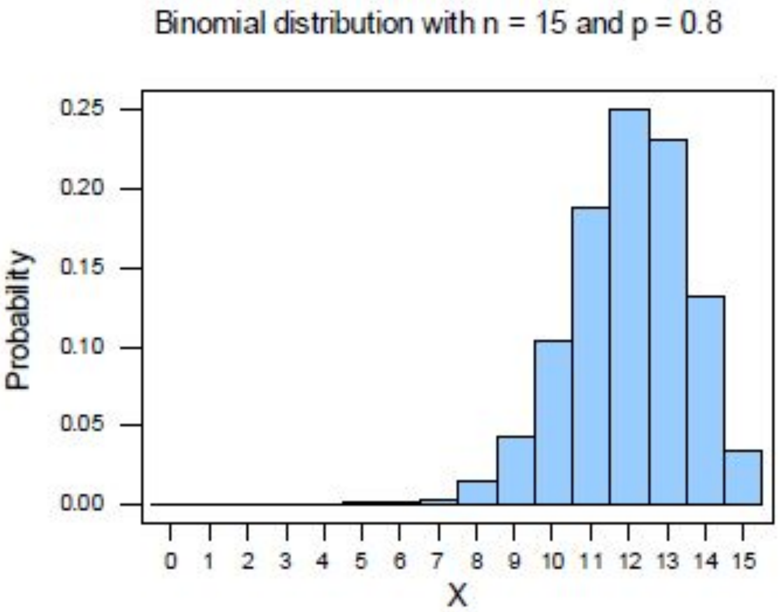
- Sample repetitively from a Bernoulli distribution n times
- Bernoulli = Binomial with n=1
- $\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$
- Expectation
  - Use linearity of expectations
  - Let X_i be the ith sample outcome from the Bernoulli distribution
  - $E[X_1 + \ldots X_n] = \sum_i E[X_i] = n \cdot p$

Binomial Distributions

PDF $\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$

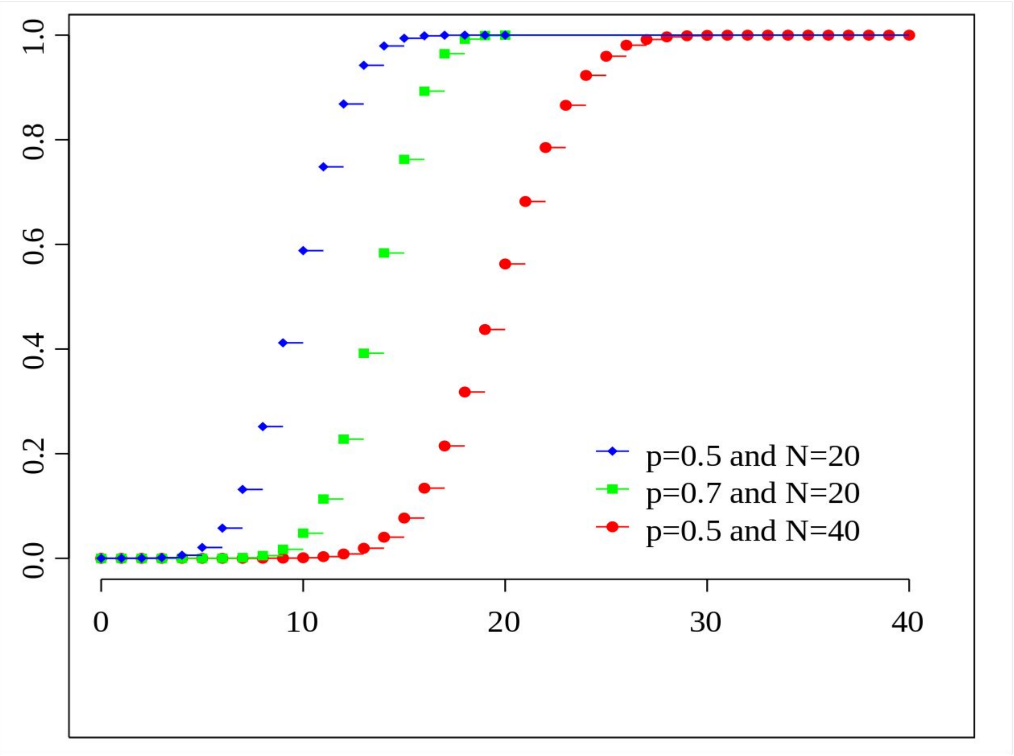






Binomial Distribution CDF

- **CDF:** $F_x(t) = \Pr(X \leq T)$



$Poisson(\lambda)$

Poisson distribution is **a probability distribution that is used to show how many times an event is likely to occur over a specified period**

- Events are independent of each other.
- The occurrence of one event does not affect the probability another event will occur.
- The average rate (events per time period) is constant.
- Two events cannot occur at the same time.

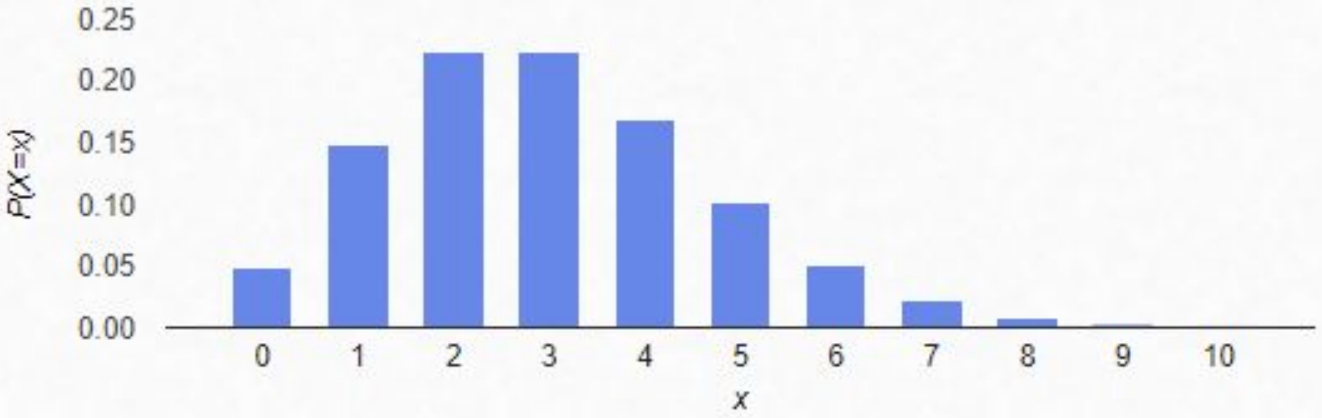
$$\Pr[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Poisson Example

- Example: You are “on call” at work. And every night between 12 and 8 AM you receive about 6 calls from customers
  - Question: What is the probability you receive a call from 2-4 AM?
  - Definitions: X~P(lambda) where X is a random variable with a Poisson distribution, and “lambda” is the mean for the interval
  - 6 calls over an 8 hour window is about ¾ of a call every hour, our window is 2 hours, so our “lambda” value is 2*3/4 = 1.5

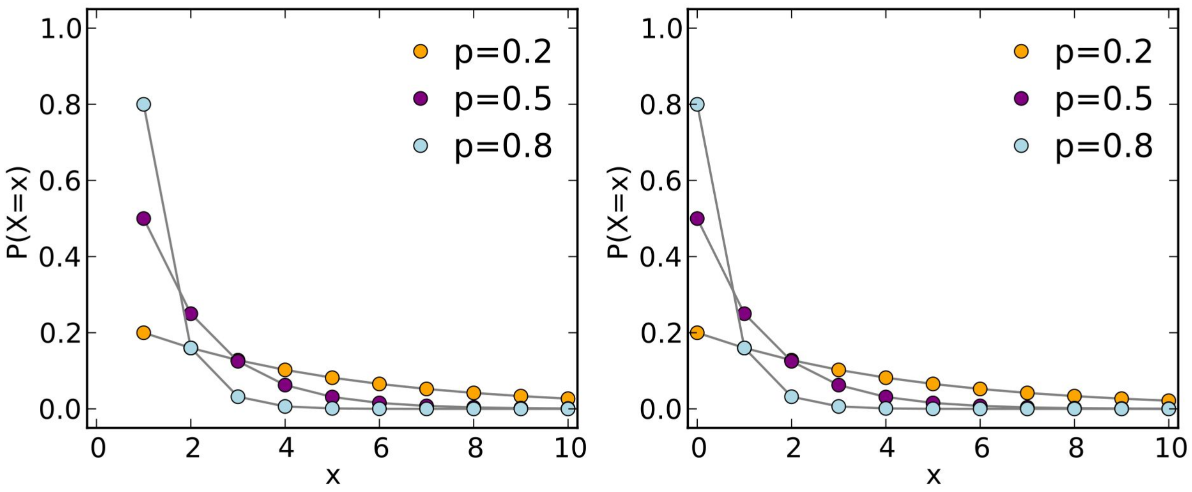
  - https://homepage.divms.uiowa.edu/~mbognar/applets/pois.html

$$\mu = E(X) = 1.5 \quad \sigma = SD(X) = 1.225 \quad \sigma^2 = Var(X) = 1.5$$



$$\mu = E(X) = 3 \quad \sigma = SD(X) = 1.732 \quad \sigma^2 = Var(X) = 3$$

Geometric(p)



- Discrete, closely related to Bernoulli
- Probability of N failures before a success
- $Pr(X = k) = (1-p)^{k-1}p^1$

- Example: What is the probability I flip a coin 4 times before getting heads?

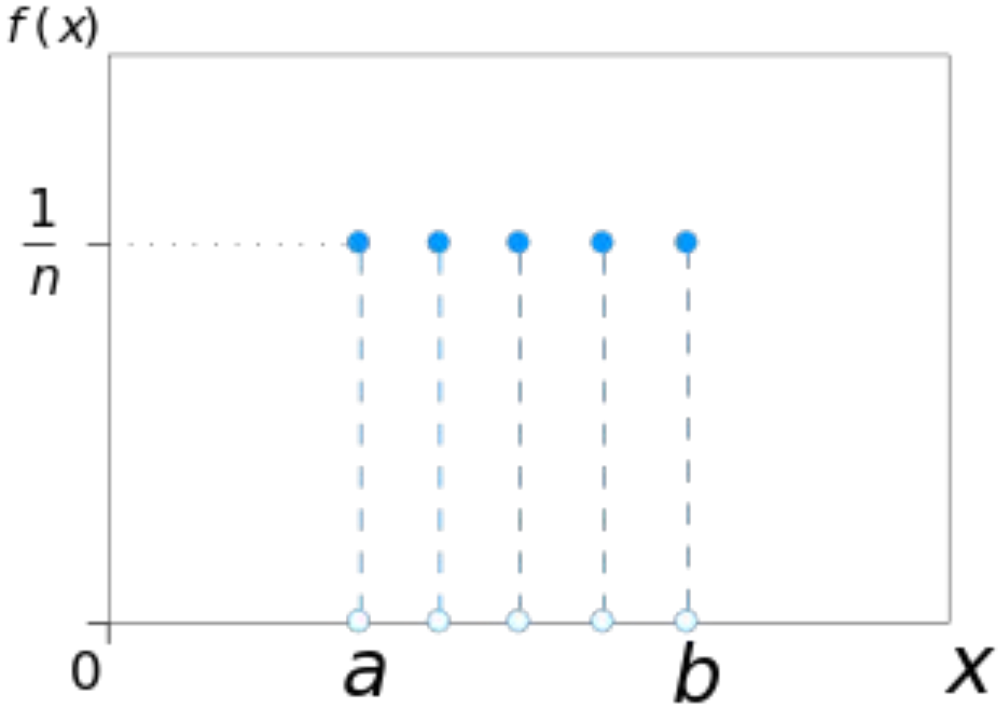
Expected Value:

$$\sum_{k=1}^{\infty} \boldsymbol{k * (1-p)^{k-1} p}$$

Geometric Expectation Derivation

1. $E[X] = \sum_{k=1}^{\infty} k * (1-p)^{k-1} p$
2. $= p * \sum_{k=1}^{\infty} \boldsymbol{k} * (\mathbf{1} - \boldsymbol{p})^{\boldsymbol{k-1}}$ // take out p from summation
3. $= p * \sum_{k=1}^{\infty} \frac{\boldsymbol{d}}{\boldsymbol{dp}} * -(\mathbf{1} - \boldsymbol{p})^{\boldsymbol{k}}$ // substitute derivative w.r.t p
4. $= p * -\frac{d}{dp} \sum_{k=1}^{\infty} (1-p)^k$ // well known series summation
5. $= -p * \frac{d}{dp} * \frac{1}{p} = -p\left(-\frac{1}{p^2}\right) = \frac{1}{p}$

Uniform - Discrete



E[X] where P(X=x) = 1/N:

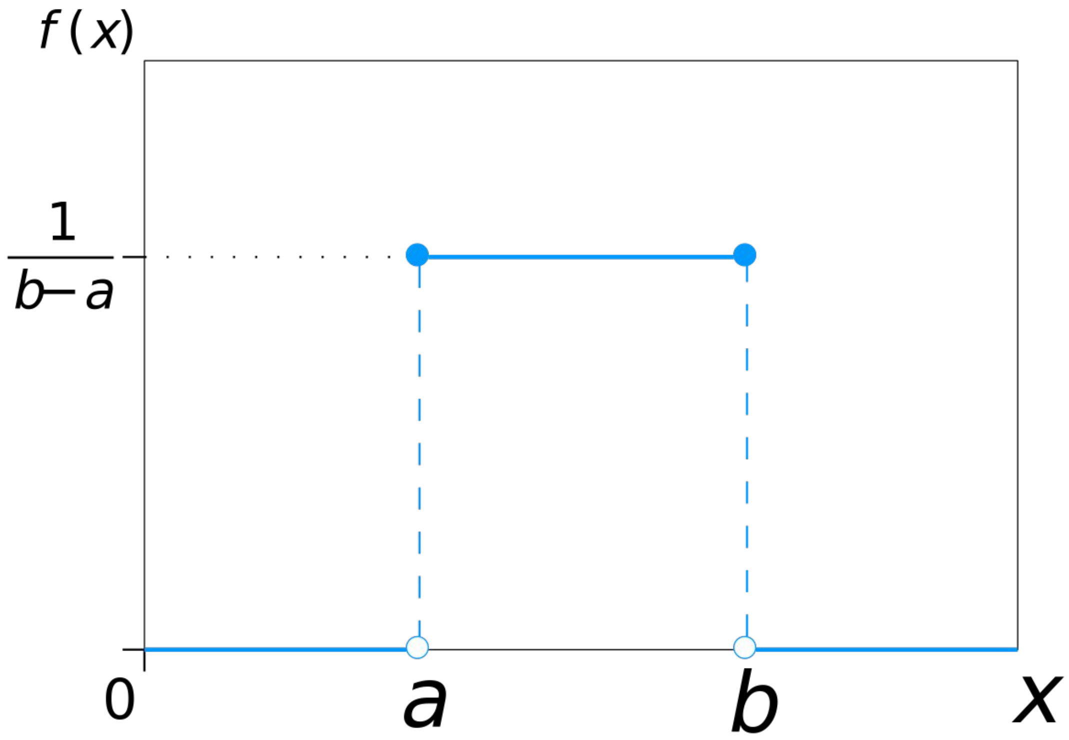
(N+1)/2

Var[x] = $\frac{N^2-1}{12}$

Continuous probability distributions

	PDF	Mean	Variance
Uniform (a,b)	$\frac{1}{b-a}$ for $a \leq x \leq b$ 0 otherwise	(a+b)/2	$(b-a)^2/12$
Exponential (λ)	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Laplace (μ, b)	$\frac{1}{2b}\exp\left(-\frac{\lvert x-\mu\rvert}{b}\right)$	μ	$2b^2$
Gaussian (μ, σ^2)	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2

Uniform - Continuous

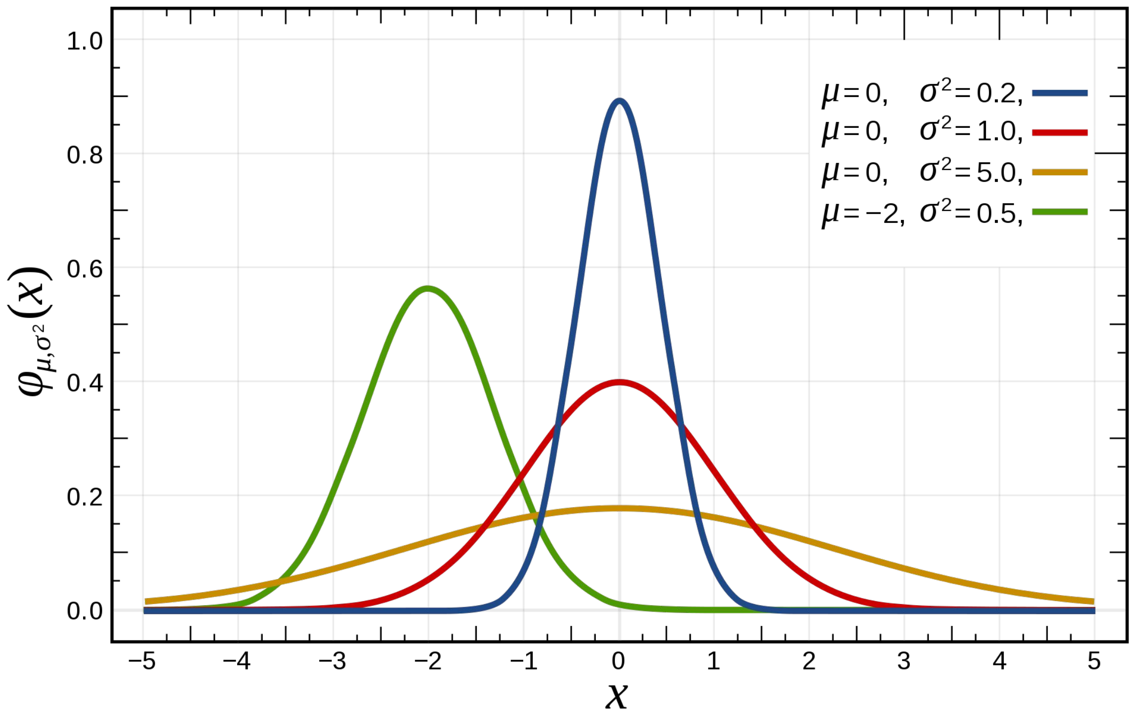


E[X] where P(X=x) = 1/N:

$$\frac{1}{b-a}$$

Var[x] = $\frac{(b-a)^2}{12}$

Gaussian – One dimension



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

- Parameterized by
  - μ = mean
  - σ = standard deviation
- It is important partly due to CLT.

Secretary problem

Imagine an administrator who wants to hire *the best* secretary out of n rankable applicants for a position.

- The applicants are interviewed one by one in random order.
- A decision about each particular applicant is to be made immediately after the interview.
- Once rejected, an applicant cannot be recalled.
- During the interview, the administrator gets a score of the current applicant, but is unaware of the quality of yet unseen applicants.

Secretary problem

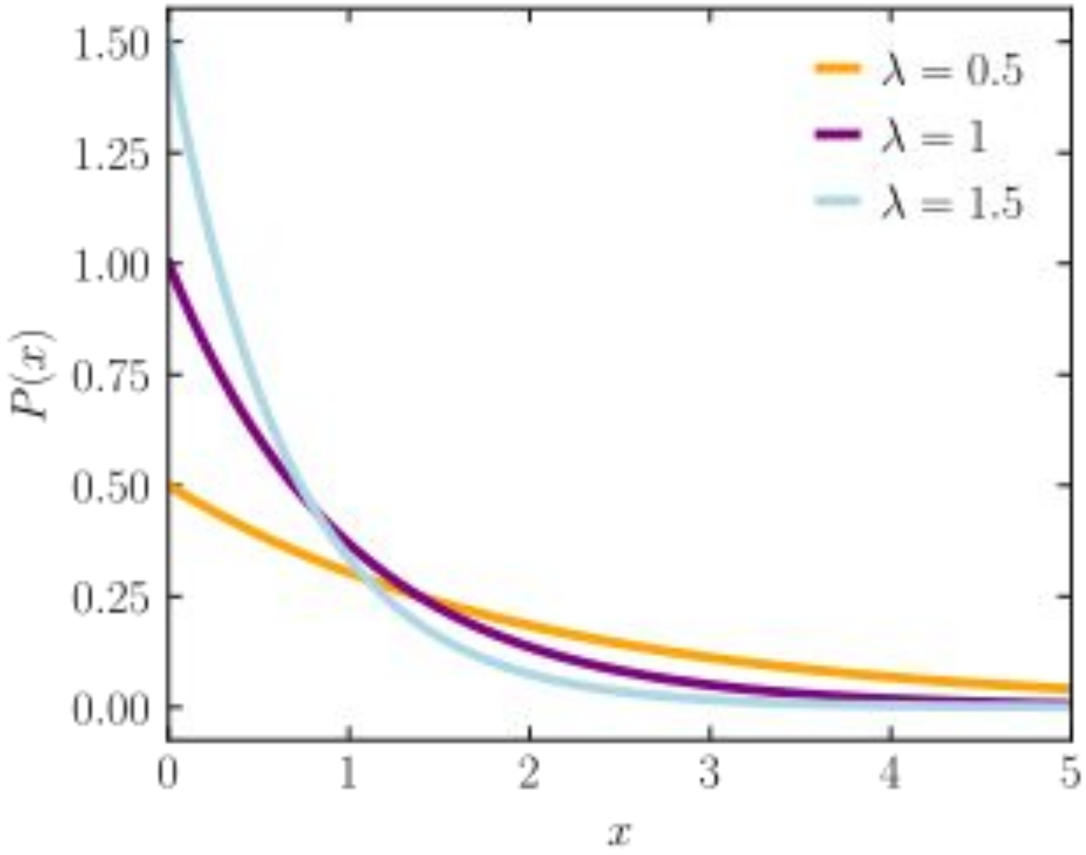
Consider the strategy: the interviewer rejects the first r − 1 applicants (let applicant M be the best applicant among these r − 1 applicants), and then selects the first subsequent applicant that is better than applicant M.

$$
\begin{aligned}
P(r) &= \sum_{i=1}^{n} P\left(\text{applicant } i \text{ is selected} \cap \text{applicant } i \text{ is the best}\right) \\
&= \sum_{i=1}^{n} P\left(\text{applicant } i \text{ is selected} | \text{applicant } i \text{ is the best}\right) \cdot P\left(\text{applicant } i \text{ is the best}\right) \\
&= \left[\sum_{i=1}^{r-1} 0 + \sum_{i=r}^{n} P\left(\begin{array}{l}\text{the best of the first } i-1 \text{ applicants} \\ \text{is in the first } r-1 \text{ applicants}\end{array} \middle| \text{applicant } i \text{ is the best}\right)\right] \cdot \frac{1}{n} \\
&= \left[\sum_{i=r}^{n} \frac{r-1}{i-1}\right] \cdot \frac{1}{n} \\
&= \frac{r-1}{n} \sum_{i=r}^{n} \frac{1}{i-1}.
\end{aligned}
$$

Optimizing this leads to r=1/e

Extra

Exponential



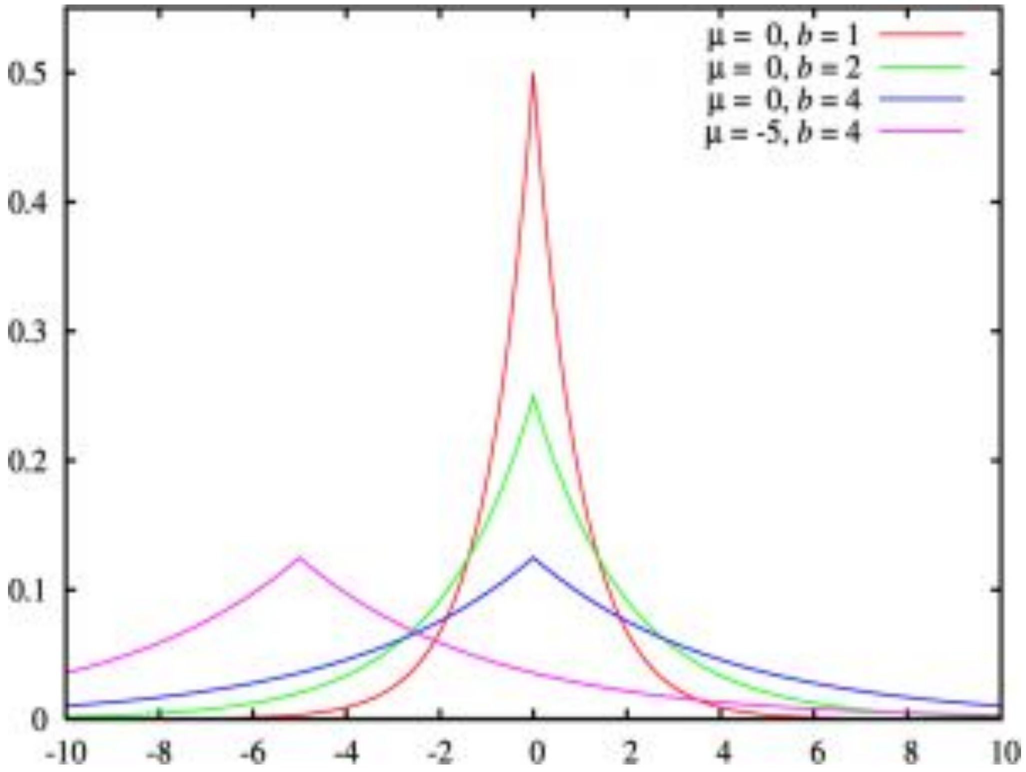
The exponential distribution is the probability distribution of the time between events in a Poisson point process.

- Parameterized by (λ)
- PDF $\Pr[X = x] = \lambda e^{-\lambda x}$

- Expected value $1/\lambda$
- Variance $1/\lambda^2$

Remark: Exponential distribution is not the same as the class of exponential families of distributions

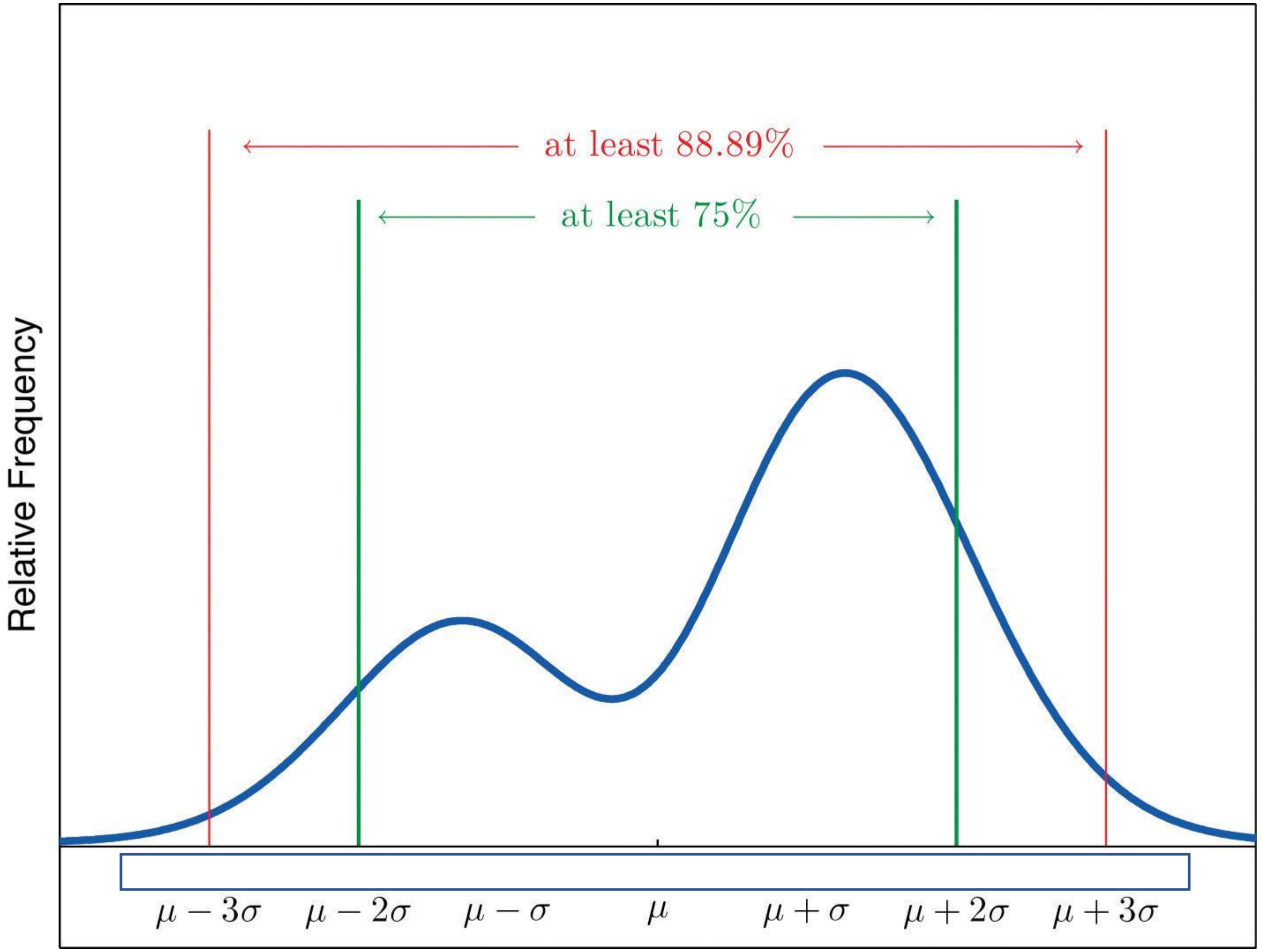
Laplace (μ, b)



- It can be thought of as two exponential distributions (with an additional location parameter) spliced together along the abscissa
- Parametrized by mu and beta:
    - Mu = Location (shift along the x axis)
    - B = scaling, maximum peak value

$$\Pr[X = x] = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

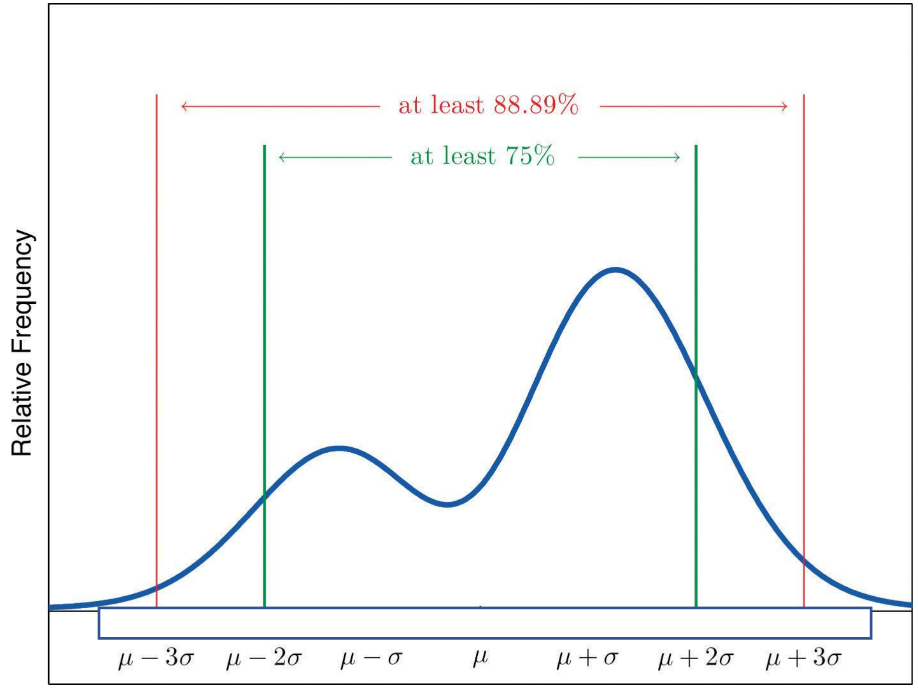
Chebyshev Inequality



- $P(|X - \mu| \geq k * \sigma) \leq \frac{1}{k^2}$

Poisson Chebyshev problem

- Let us say that X is modeled by a Poisson variable with mu/lambda/mean = 100
- What is the lower bound of P(75 < X < 125)?



Poisson Chebyshev solution

- Find μ and σ:
  - μ = 100
  - $\sigma = \sqrt{var} = \sqrt{100} = 10$

- $P(|X - \mu| \geq k * \sigma) \leq \frac{1}{k^2}$ is the P that we exist outside a certain SD
- What we want is the P we exist WITHIN a certain SD
- P(75 < X < 125) $=$ 1 - $P(|X - 100| \geq 2.5 * 10) \geq 1 - \frac{1}{2.5^2} = 0.84$

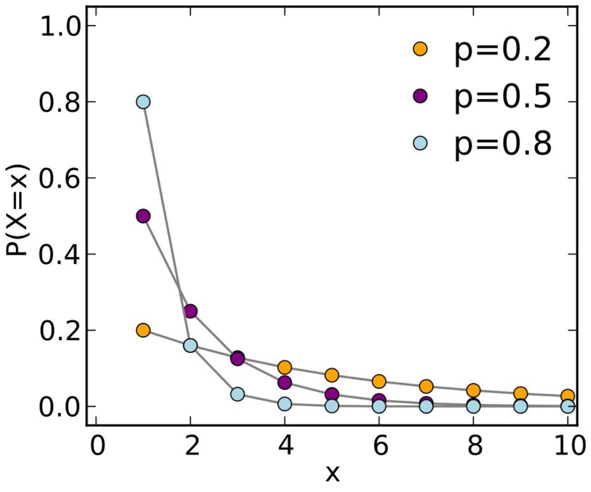
Practice Problem

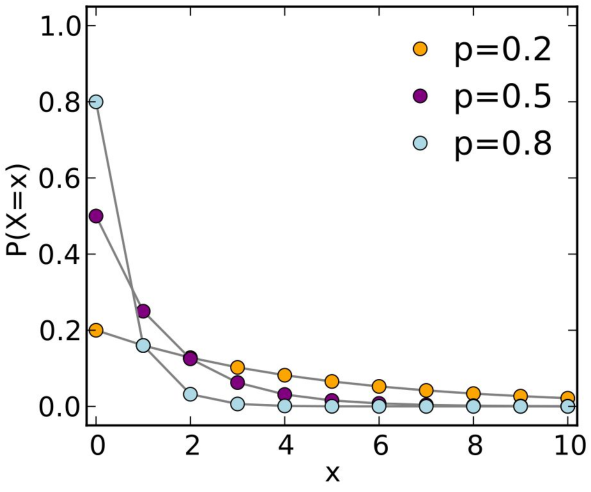
- Broken device, calling customer support
- 1% chance a human representative being available and picking up my call, or else robot will disconnect me

- Question:
  - What is the expected number of times I will have to call in before I talk to a human being?

Practice Solution

- Identify an analytical distribution:
  - Geometric in our case (fails until a success)





- Discrete, closely related to Bernoulli
- Probability of N failures before a success
- $P(X = k) = (1-p)^{k-1}p^1$
- Example: What is the probability I flip a coin 4 times before getting heads?

Practice Solution

- X = number of call attempts before a human picks up
- p = .01, q = 1-p
- $E[X] = \sum_{i=0}^{N} X_i * P(X_i)$
- $= 1(.01) + 2(.99)^1(.01) + 3(.99)^2(.01) + \ldots$
- = $\mathrm{p}/(1-q)^2$ = 1/p = 1/.01 = **100**