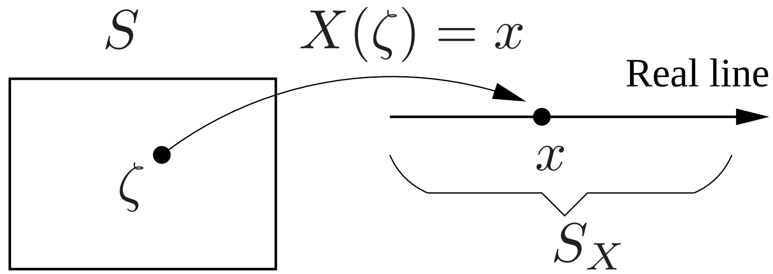
2. Random Variables

- definition
- probability measures: CDF, PMF, PDF
- expected values and moments
- examples of RVs

Definition



a random variable X is a *function* mapping an outcome to a real number

- the sample space, S, is the *domain* of the random variable
- S_X is the range of the random variable

example: toss a coin three times and note the sequence of heads and tails

$$S = \{\text{HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}\}$$

Let X be the number of heads in the three tosses

$$S_X = \{0, 1, 2, 3\}$$

Types of Random Variables

Discrete RVs take values from a countable set

example: let X be the number of times a message needs to be transmitted until it arrives correctly

$$S_X = \{1, 2, 3, \ldots\}$$

Continuous RVs take an infinite number of possible values

example: let X be the time it takes before receiving the next phone calls

Mixed RVs have some part taking values over an interval like typical continuous variables, and part of it concentrated on particular values like discrete variables

Probability measures

Cumulative distribution function (CDF)

$$F(a) = P(X \leq a)$$

Probability mass function (PMF) for discrete RVs

$$p(k) = P(X = k)$$

Probability density function (PDF) for continuous RVs

$$f(x) = \frac{dF(x)}{dx}$$

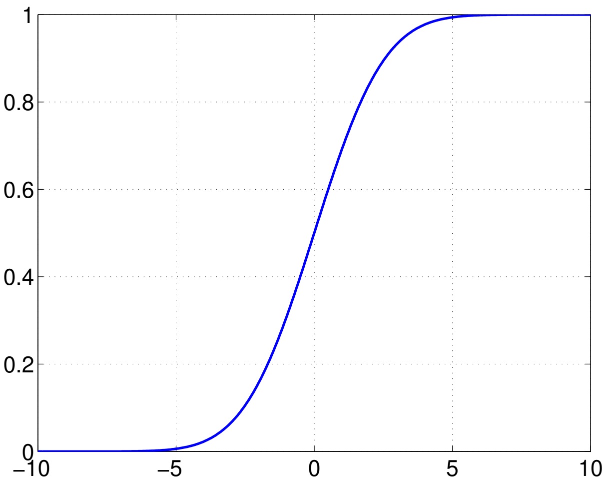
Cumulative Distribution Function (CDF)

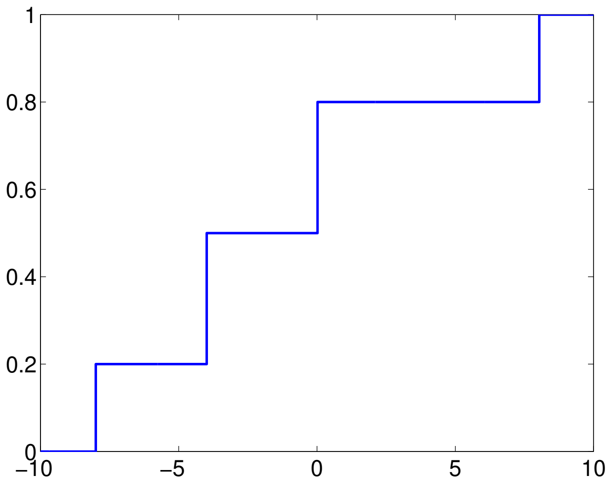
Properties

$$0 \le F(a) \le 1$$

$$F(a) \to 1, \quad \text{as } a \to \infty$$

$$F(a) \to 0, \quad \text{as } a \to -\infty$$





$$F(b) - F(a) = \int_a^b f(x)dx \qquad\qquad F(a) = \sum_{k \le a} p(k)$$

Probability Density Function

Probability Density Function (PDF)

- $f(x) \geq 0$
- $P(a \leq X \leq b) = \int_a^b f(x)dx$
- $F(x) = \int\limits_{-\infty}^{x} f(u)du$

Probability Mass Function (PMF)

- $p(k) \geq 0$ for all k
- $\sum\limits_{k \in S} p(k) = 1$

Expected values

let $g(X)$ be a function of random variable X

$$\mathbf{E}[g(X)] = \begin{cases} \sum\limits_{x \in S} g(x)p(x) & X \text{ is discrete} \\ \int\limits_{-\infty}^{\infty} g(x)f(x)dx & X \text{ is continuous} \end{cases}$$

Mean

$$\mu = \mathbf{E}[X] = \begin{cases} \sum\limits_{x \in S} xp(x) & X \text{ is discrete} \\ \int\limits_{-\infty}^{\infty} xf(x)dx & X \text{ is continuous} \end{cases}$$

Variance

$$\sigma^2 = \mathbf{var}[X] = \mathbf{E}[(X-\mu)^2]$$

n^{th} **Moment**

$$\mathbf{E}[X^n]$$

Facts

Let $Y = g(X) = aX + b$, $\quad a, b$ are constants

- $\mathbf{E}[Y] = a\mathbf{E}[X] + b$
- $\mathbf{var}[Y] = a^2\,\mathbf{var}[X]$
- $\mathbf{var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$

Example of Random Variables

Discrete RVs

- Bernoulli
- Binomial
- Multinomial
- Geometric
- Negative binomial
- Poisson
- Uniform

Continuous RVs

- Uniform
- Exponential
- Gaussian (Normal)
- Gamma
- Rayleigh
- Cauchy
- Laplacian

Bernoulli random variables

let A be an event of interest

a Bernoulli random variable X is defined as

$$X = 1 \text{ if } A \text{ occurs} \quad \text{and} \quad X = 0 \text{ otherwise}$$

it can also be given by the *indicator function* for A

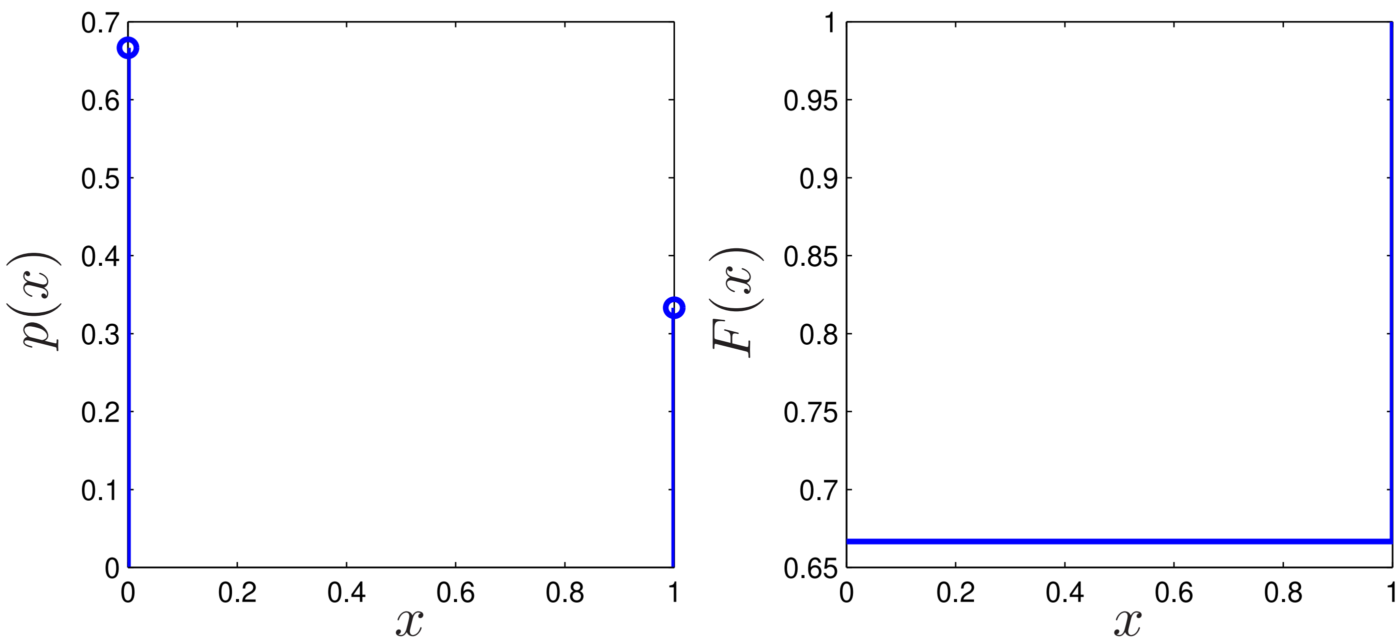
$$X(\zeta) = \begin{cases} 0, & \text{if } \zeta \text{ not in } A \\ 1, & \text{if } \zeta \text{ in } A \end{cases}$$

PMF: $p(1) = p, \quad p(0) = 1 - p, \quad 0 \leq p \leq 1$

Mean: $\mathbf{E}[X] = p$

Variance: $\mathbf{var}[X] = p(1-p)$

Example of Bernoulli PMF: $p = 1/3$



Binomial random variables

- X is the number of successes in a sequence of n independent trials
- each experiment yields success with probability p
- when $n = 1$, X is a Bernoulli random variable
- $S_X = \{0, 1, 2, \ldots, n\}$
- ex. Transmission errors in a binary channel: X is the number of errors in n independent transmissions

PMF

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

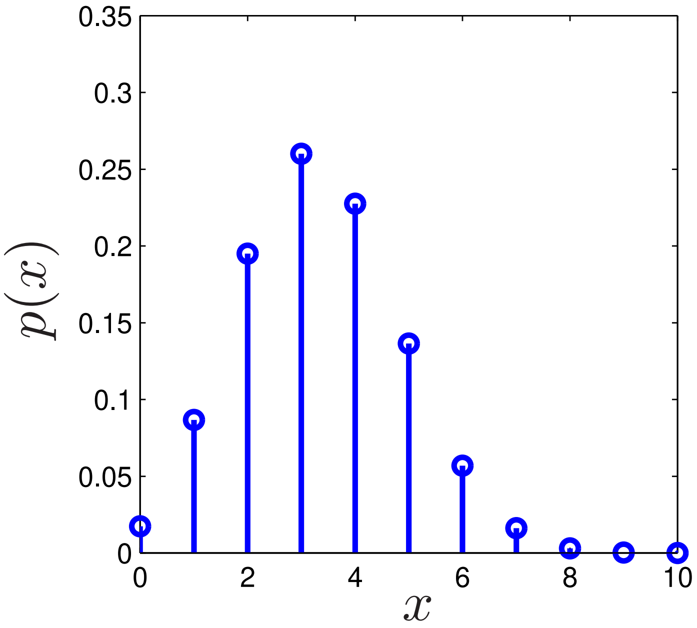
Mean

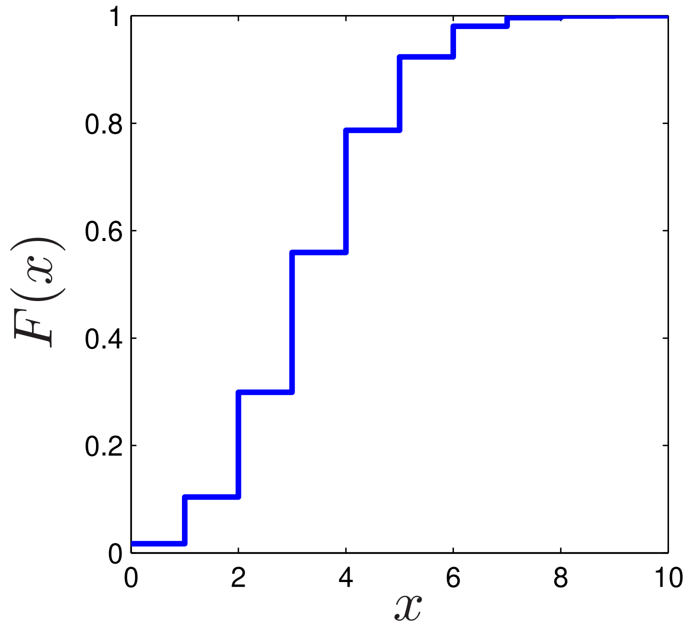
$$\mathbf{E}[X] = np$$

Variance

$$\mathbf{var}[X] = np(1-p)$$

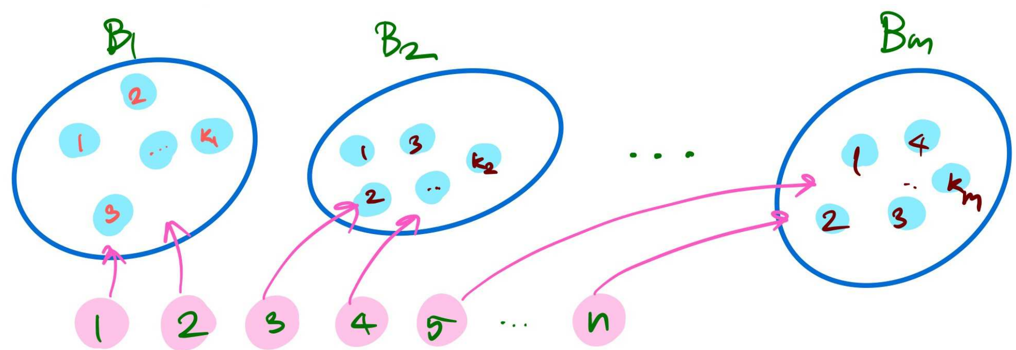
Example of Binomial PMF: $p = 1/3, n = 10$





Multinomial coefficient

suppose we partition a set of n objects into m subsets $B_1, B_2, \ldots, B_m$



- B_i is assigned k_i elements and $k_1 + k_2 + \cdots + k_m = n$
- denote N_i the number of possible assignments to the subset B_i

$$N_1 = \binom{n}{k_1}, N_2 = \binom{n-k_1}{k_2}, \ldots, N_{m-1} = \binom{n - k_1 - k_2 - \cdots - k_{m-2}}{k_{m-1}}$$

- the number of possible partitions is $N_1 N_2 \cdots N_{m-1} = \frac{n!}{k_1! k_2! \cdots k_m!}$ and is called the **multinomial coefficient**

Mutinomial random variables

- a generalization of binomial random variables to consider a trial having more than two possible outcomes
- in each trial, there are m possible events, denoted by $B_1, B_2, \ldots, B_m$ with

$$P(B_k) = p_k, \quad \text{and} \quad p_1 + p_2 + \cdots + p_m = 1$$

- suppose n independent repetitions of the experiment are performed
- let X_j be the number of times each B_j occurs

$$P(X_1 = k_1, X_2 = k_2, \ldots, X_m = k_m) = \frac{n!}{k_1!k_2!\ldots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$$

  where $k_1 + k_2 + \cdots + k_m = n$

- the multinomial coefficient is the number of possible orderings that $X_1 = k, \ldots, X_m = k_m$

PMF: the joint probability of vector $X = (X_1, X_2, \ldots, X_m)$

$$P(X_1 = k_1, X_2 = k_2, \ldots, X_m = k_m) = \frac{n!}{k_1!k_2!\ldots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$$

where $k_i \in \{0, 1, \ldots, n\}$ and $k_1 + k_2 + \cdots + k_m = 1$

Mean

$$\mathbf{E}[X_i] = np_i$$

Variance

$$\mathbf{var}[X_i] = np_i(1 - p_i), \quad \mathbf{cov}(X_i, X_j) = -np_ip_j, \;\; i \neq j$$

some applications:

- the data of N samples can be categorized into K classes, e.g., N subjects with blood types of A, B, AB, and O
- multinomial logistic regression in K-class classification

Geometric random variables

- repeat independent Bernoulli trials, each has probability of success p
- X is the number of experiments required until the first success occurs
- $S_X = \{1, 2, 3, \ldots\}$
- ex. Message transmissions: X is the number of times a message needs to be transmitted until it arrives correctly

PMF

$$p(k) = P(X = k) = (1-p)^{k-1}p$$

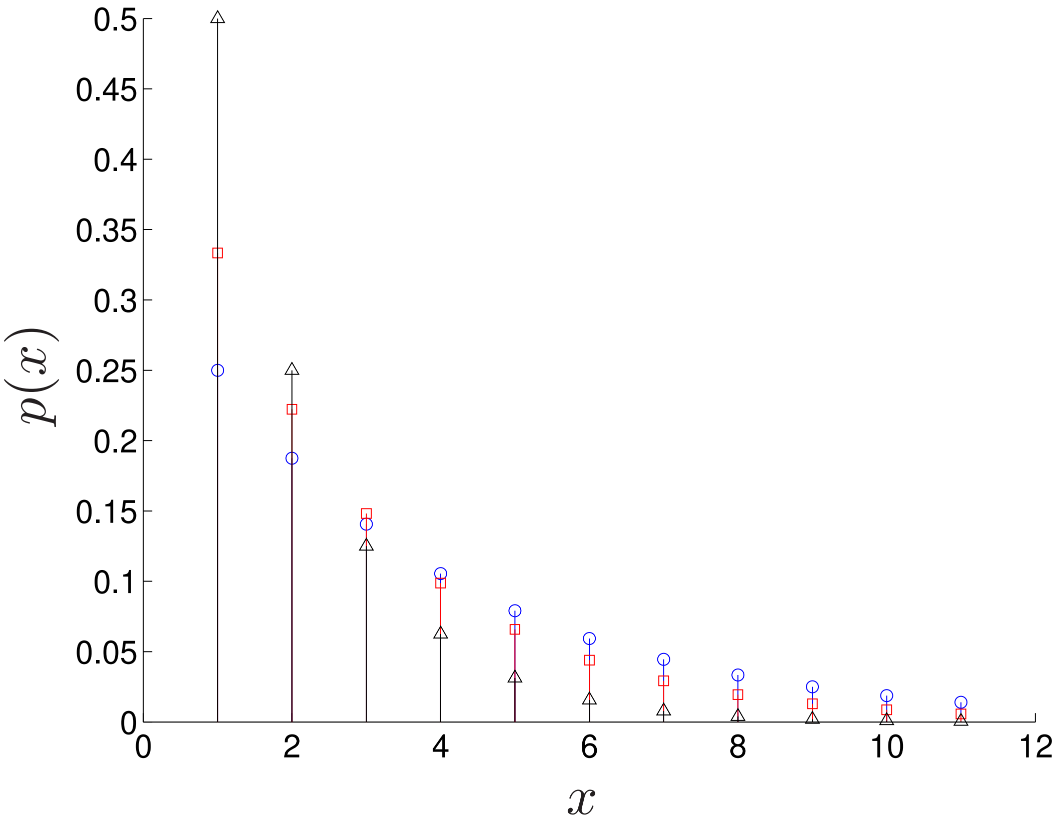
Mean

$$\mathbf{E}[X] = \frac{1}{p}$$

Variance

$$\mathbf{var}[X] = \frac{1-p}{p^2}$$

Example of Geometric PMF: $p = 1/3$



- parameters: $p = 1/4, 1/3, 1/2$

Negative binomial (Pascal) random variables

- repeat independent Bernoulli trials until observing the r^{th} success
- X is the number of trials required until the r^{th} success occurs
- X can be viewed as the sum of r geometrically RVs
- $S_X = \{r, r+1, r+2, \ldots\}$

PMF

$$p(k) = P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \ldots$$

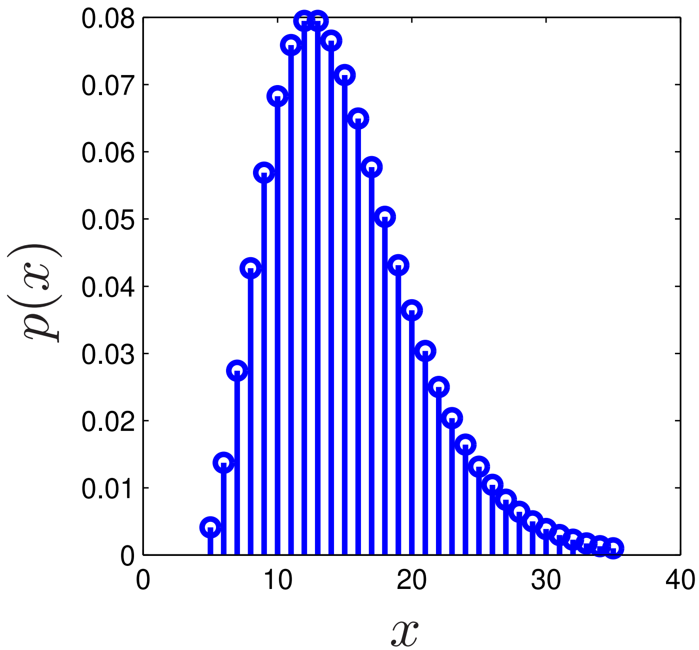
Mean

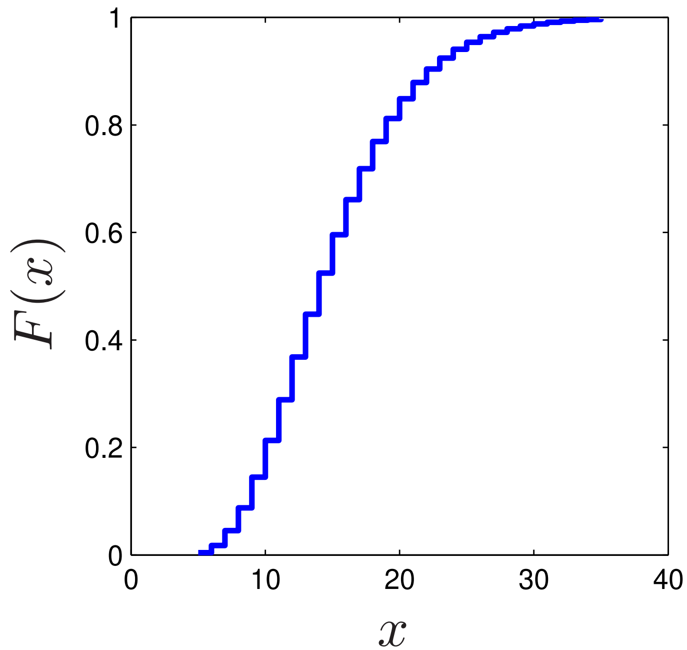
$$\mathbf{E}[X] = \frac{r}{p}$$

Variance

$$\mathbf{var}[X] = \frac{r(1-p)}{p^2}$$

Example of negative binomial PMF: $p = 1/3, r = 5$





Poisson random variables

- X is a number of events occurring in a certain period of time
- events occur with a known average rate
- the expected number of occurrences in the interval is λ
- $S_X = \{0, 1, 2, \ldots\}$
- examples:
  - number of emissions of a radioactive mass during a time interval
  - number of queries arriving in t seconds at a call center
  - number of packet arrivals in t seconds at a multiplexer

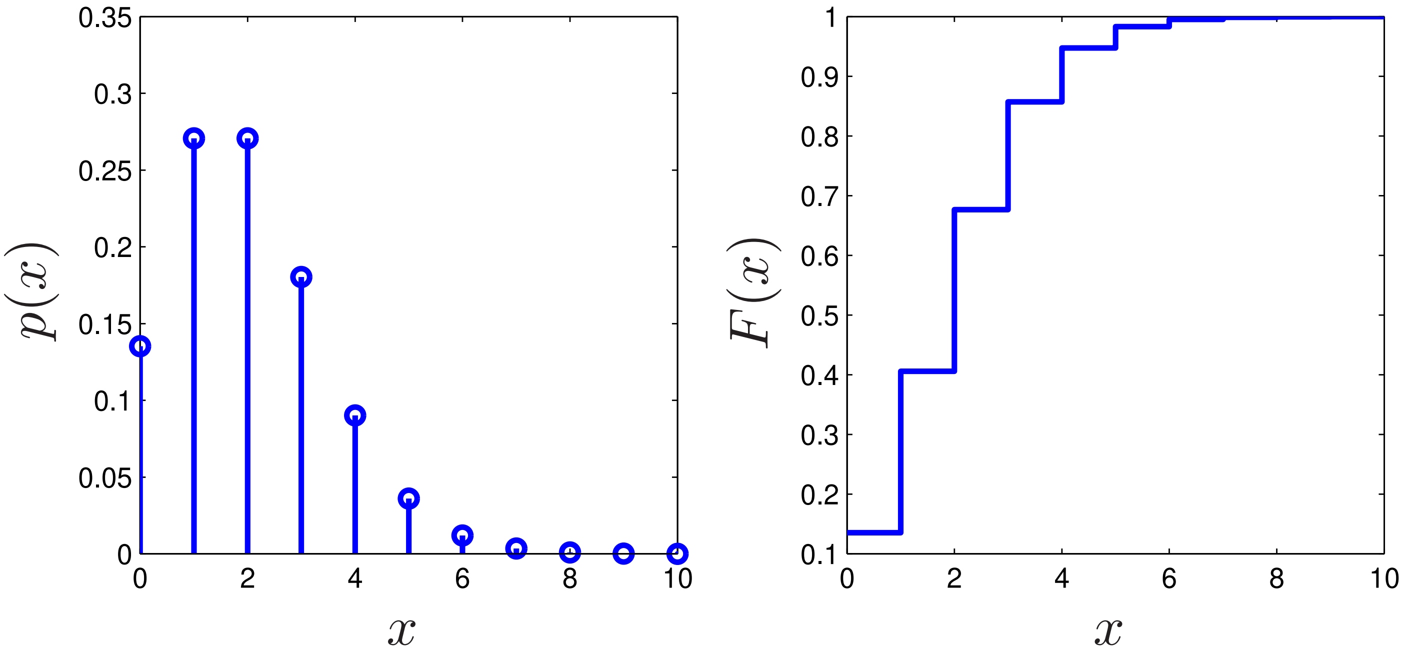
PMF

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

Mean $\mathbf{E}[X] = \lambda$

Variance $\mathbf{var}[X] = \lambda$

Example of Poisson PMF: $\lambda = 2$



Derivation of Poisson distribution

- approximate a binomial RV when n is large and p is small
- define $\lambda = np$, in 1898 Bortkiewicz showed that

$$p(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

Proof.

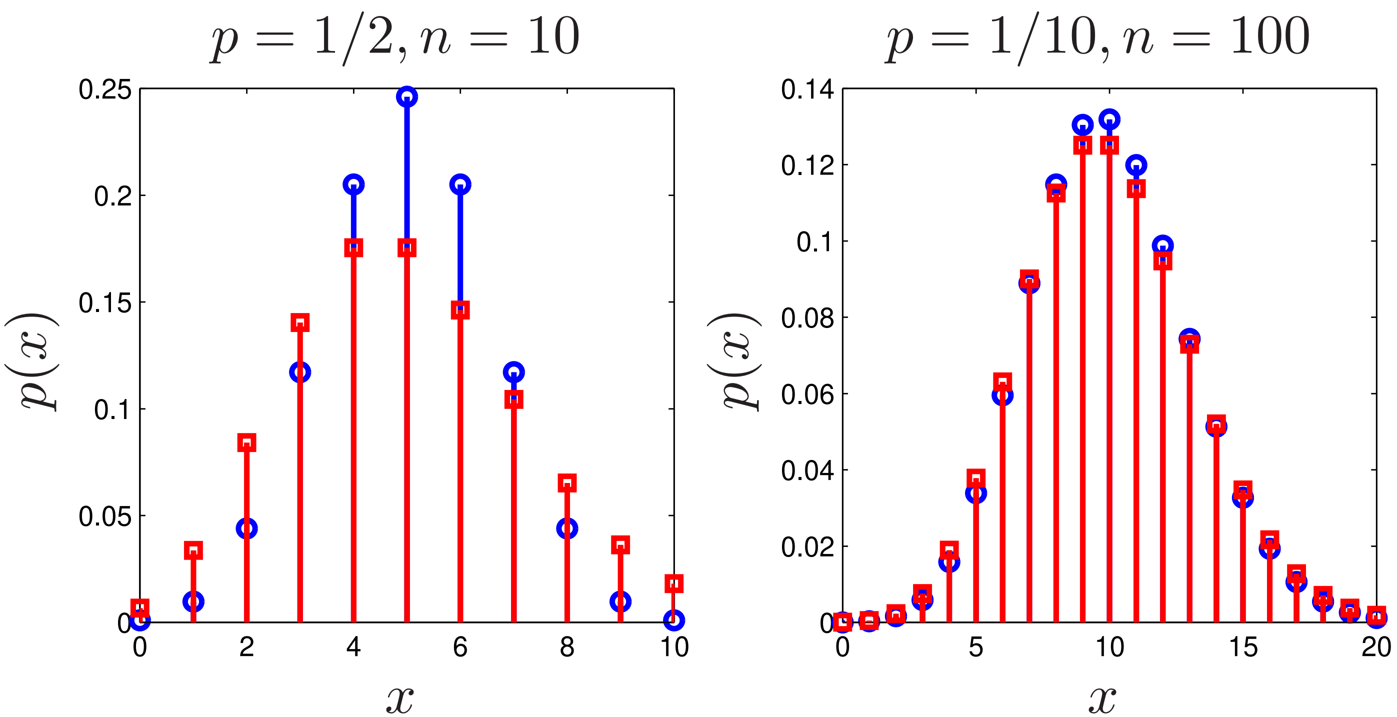
$$p(0) = (1-p)^n = (1-\lambda/n)^n \approx e^{-\lambda}, \quad n \rightarrow \infty$$

$$\frac{p(k+1)}{p(k)} = \frac{(n-k)p}{(k+1)(1-p)} = \frac{(1-k/n)\lambda}{(k+1)(1-\lambda/n)}$$

take the limit $n \rightarrow \infty$

$$p(k+1) = \frac{\lambda}{k+1} p(k) = \left(\frac{\lambda}{k+1}\right) \left(\frac{\lambda}{k}\right) \cdots \left(\frac{\lambda}{1}\right) p(0) = \frac{\lambda^{k+1}}{(k+1)!} e^{-\lambda}$$

Comparison of Poisson and Binomial PMFs



- red: Poisson
- blue: Binomial

Exponential random variables

- arise when describing the time between occurrence of events
- examples:
  - the time between customer demands for call connections
  - the time used for a bank teller to serve a customer
- λ is the rate at which events occur
- a continuous counterpart of the geometric random variable

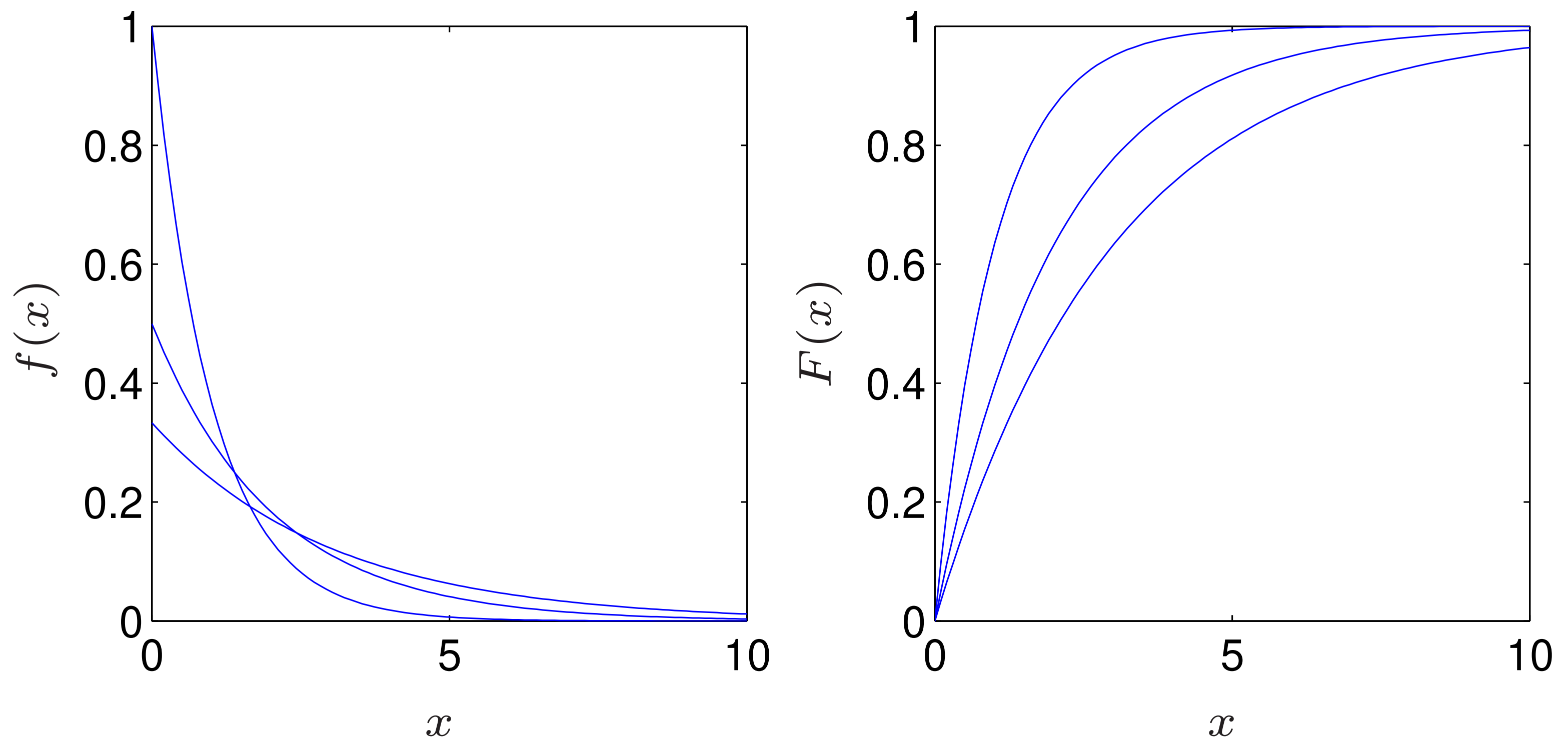
PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

Mean $\mathbf{E}[X] = \frac{1}{\lambda}$

Variance $\mathbf{var}[X] = \frac{1}{\lambda^2}$

Example of Exponential PDF



- parameters: $\lambda = 1, 1/2, 1/3$

Memoryless property

$$P(X > t + h | X > t) = P(X > h)$$

- $P(X > t + h | X > t)$ is the probability of having to wait additionally at least h seconds given that one has already been waiting t seconds
- $P(X > h)$ is the probability of waiting at least h seconds when one first begins to wait
- thus, the probability of waiting at least an additional h seconds is the same regardless of how long one has already been waiting

Proof.

$$\begin{aligned} P(X > t + h | X > t) &= \frac{P\{(X > t + h) \cap (X > t)\}}{P(X > t)}, \quad \text{for } h > 0 \\ &= \frac{P(X > t + h)}{P(X > t)} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} \end{aligned}$$

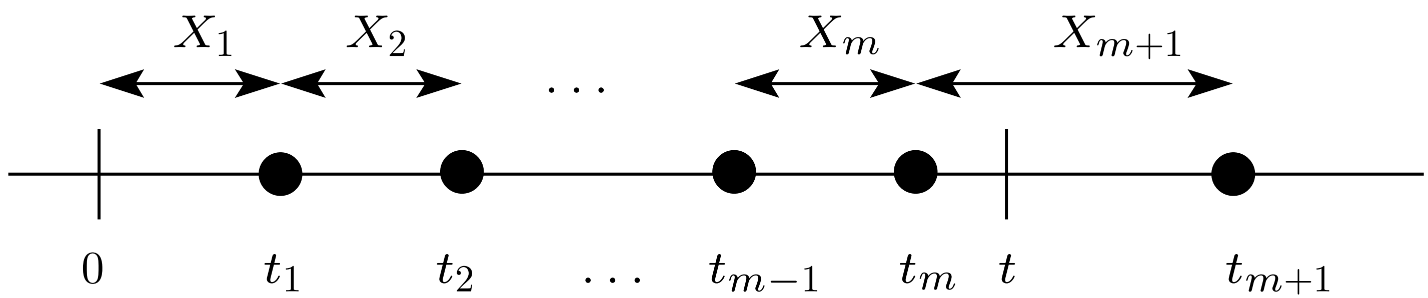
this is not the case for other non-negative continuous RVs

in fact, the conditional probability

$$P(X > t + h | X > t) = \frac{1 - P(X \leq t + h)}{1 - P(X \leq t)} = \frac{1 - F(t + h)}{1 - F(t)}$$

depends on t in general

m-Erlang random variables



- the kth event occurs at time t_k
- the times $X_1, X_2, \ldots, X_m$ between events are exponential RVs
- $N(t)$ denotes the number of events in t seconds, which is a Poisson RV
- $S_m = X_1 + X_2 + \cdots + X_m$ is the elapsed time until the mth occurs

we can show that S_m is an m-Erlang random variable

Proof. $S_m \le t$ iff m or more events occur in t seconds

$$\begin{aligned} F(t) &= P(S_m \le t) = P(N(t) \ge m) \\ &= 1 - \sum_{k=0}^{m-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

to get the density function of S_m, we take the derivative of $F(t)$:

$$\begin{aligned} f(t) &= \frac{dF(t)}{dt} = \sum_{k=0}^{m-1} \frac{e^{-\lambda t}}{k!} \left(\lambda(\lambda t)^k - k\lambda(\lambda t)^{k-1}\right) \\ &= \frac{\lambda(\lambda t)^{m-1} e^{-\lambda t}}{(m-1)!} \quad \Rightarrow \quad \text{Erlang distribution with parameters } m, \lambda \end{aligned}$$

- the sum of m exponential RVs with rate λ is an m-Erlang RV
- if m becomes large, the m-Erlang RV should approach the normal RV
- from the pdf, m-erlang is a special case of gamma variable with parameter $\alpha = m$

Uniform random variables

Discrete Uniform RVs

- X has n possible values, $x_1, \dots, x_n$ that are equally probable
- **PMF**

$$p(x) = \begin{cases} \frac{1}{n}, & \text{if } x \in \{x_1, \dots, x_n\} \\ 0, & \text{otherwise} \end{cases}$$

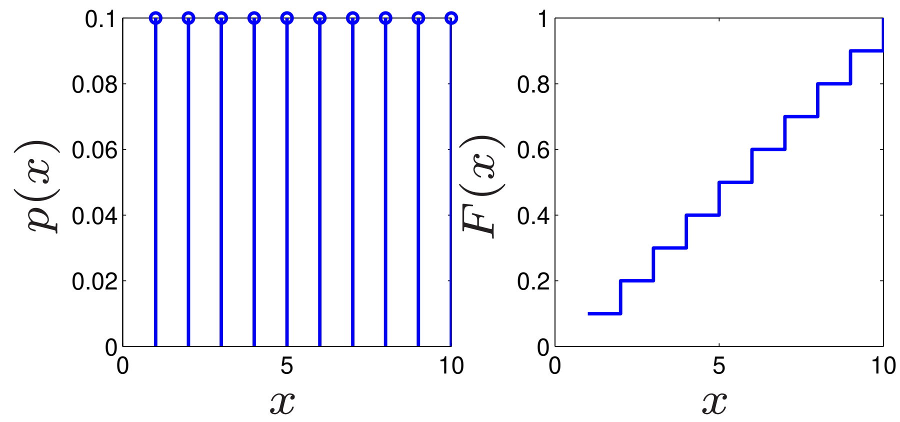
Continuous Uniform RVs

- X takes any values on an interval $[a, b]$ that are equally probable
- **PDF**

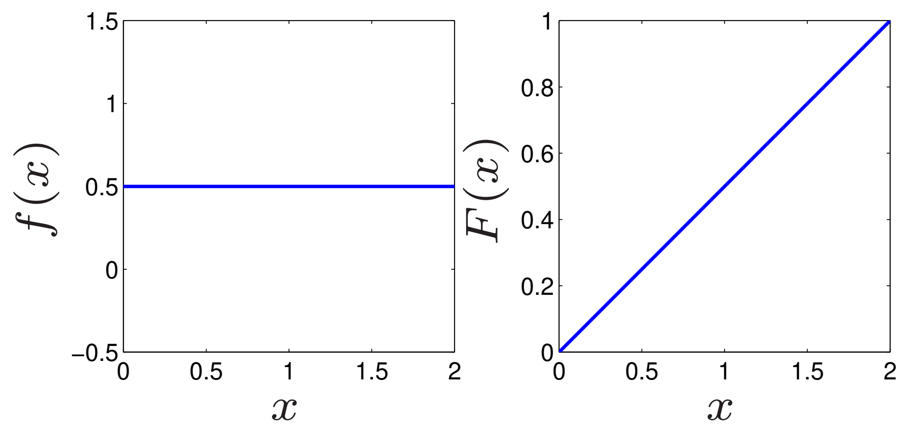
$$f(x) = \begin{cases} \frac{1}{(b-a)}, & \text{for } x \in [a, b] \\ 0, & \text{otherwise} \end{cases}$$

- **Mean:** $\mathbf{E}[X] = (a + b)/2$
- **Variance:** $\mathbf{var}[X] = (b - a)^2/12$

Example of Discrete Uniform PMF: $X = 0, 1, 2, \ldots, 10$



Example of Continuous Uniform PMF: $X \in [0, 2]$



Gaussian (Normal) random variables

- arise as the outcome of the *central limit theorem*
- the sum of a *large* number of RVs is distributed approximately normally
- many results involving Gaussian RVs can be derived in analytical form
- let X be a Gaussian RV with parameters mean μ and variance σ^2

Notation $$X \sim \mathcal{N}(\mu, \sigma^2)$$

PDF

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \,\mathsf{exp} - \frac{(x-\mu)^2}{2\sigma^2}, \quad -\infty < x < \infty$$

Mean $\mathbf{E}[X] = \mu$

Variance $\mathbf{var}[X] = \sigma^2$

let $Z \sim \mathcal{N}(0,1)$ be the normalized Gaussian variable

CDF of Z is

$$F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt \;\; \triangleq \;\; \Phi(z)$$

then CDF of $X \sim \mathcal{N}(\mu, \sigma^2)$ can be obtained by

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

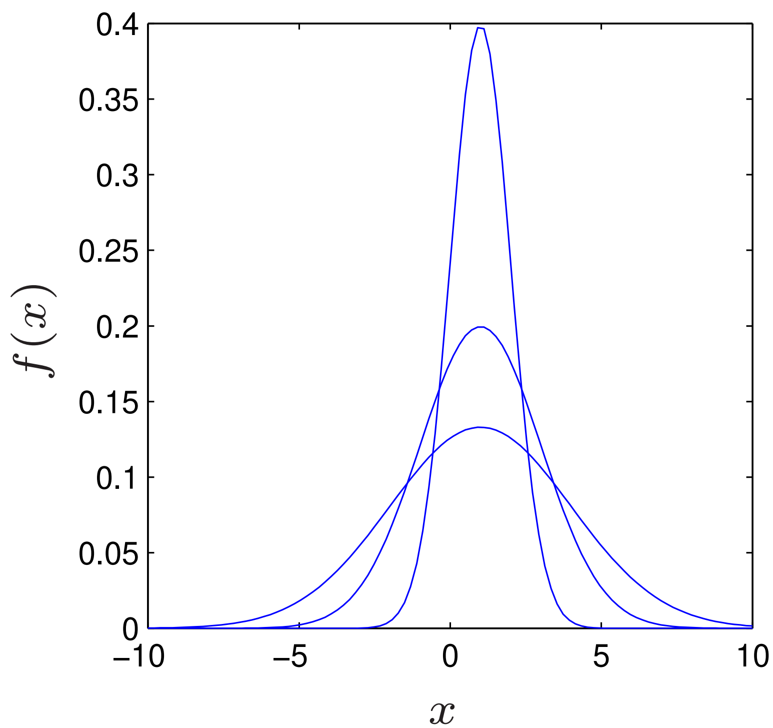
in MATLAB, the error function is defined as

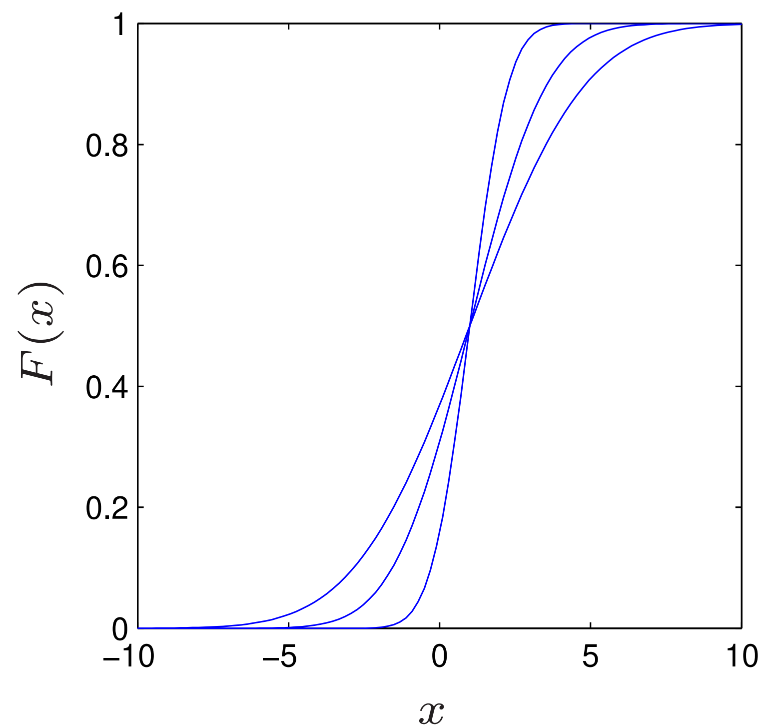
$$\mathsf{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

hence, $\Phi(z)$ can be computed via the `erf` command as

$$\Phi(z) = \frac{1}{2}\left[1 + \mathsf{erf}\left(\frac{z}{\sqrt{2}}\right)\right]$$

Example of Gaussian PDF





- parameters: $\mu = 1$, $\sigma = 1, 2, 3$

Gamma random variables

- appears in many applications:
  - the time required to service customers in queuing system
  - the lifetime of devices in reliability studies
  - the defect clustering behavior in VLSI chips
- let X be a Gamma variable with parameters α, λ

PDF

$$f(x) = \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}, \quad x \geq 0; \quad \alpha, \lambda > 0$$

where $\Gamma(z)$ is the gamma function, defined by

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx, \quad z > 0$$

Mean $\mathbf{E}[X] = \frac{\alpha}{\lambda}$ **Variance** $\mathbf{var}[X] = \frac{\alpha}{\lambda^2}$

Properties of the gamma function

$$
\begin{aligned}
\Gamma(1/2) &= \sqrt{\pi} \\
\Gamma(z+1) &= z\Gamma(z) \quad \text{for } z > 0 \\
\Gamma(m+1) &= m!, \quad \text{for } m \text{ a nonnegative integer}
\end{aligned}
$$

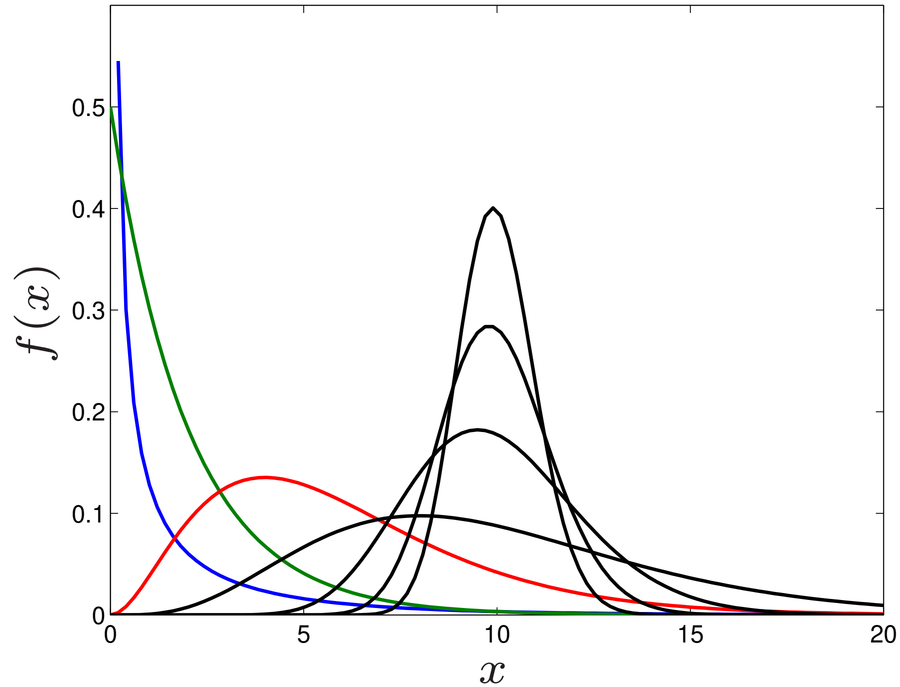
the value of $\Gamma(1/2)$ is obtaind by a change of variable $u = \sqrt{x}$ to Gaussian

Special cases

a Gamma RV becomes

- exponential RV when $\alpha = 1$
- m-Erlang RV when $\alpha = m$, a positive integer
- chi-square RV with k DOF when $\alpha = k/2, \lambda = 1/2$

Example of Gamma PDF



- blue: $\alpha = 0.2, \lambda = 0.2$ (long tail)
- green: $\alpha = 1, \lambda = 0.5$ (exponential)
- red: $\alpha = 3, \lambda = 1/2$ (Chi square with 6 DOF)
- black: $\alpha = 5, 20, 50, 100$ and $\alpha/\lambda = 10$ (α-Erlang with mean 10)

Chi-squared random variables

- arise as a sum of k i.i.d. Gaussian variables
- ex. sample variance of i.i.d. Gassian samples $\{X_1, \ldots, X_N\}$ with variance σ^2; it is well-known that $(N-1)s^2/\sigma^2$ is $\mathcal{X}^2_{N-1}$
- appear in asymptotic properties of estimators
- $X \sim \mathcal{X}^2_k$: chi-square variable with degree of freedom k

PDF

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad x \geq 0, \quad k \in \mathbf{Z}^+$$

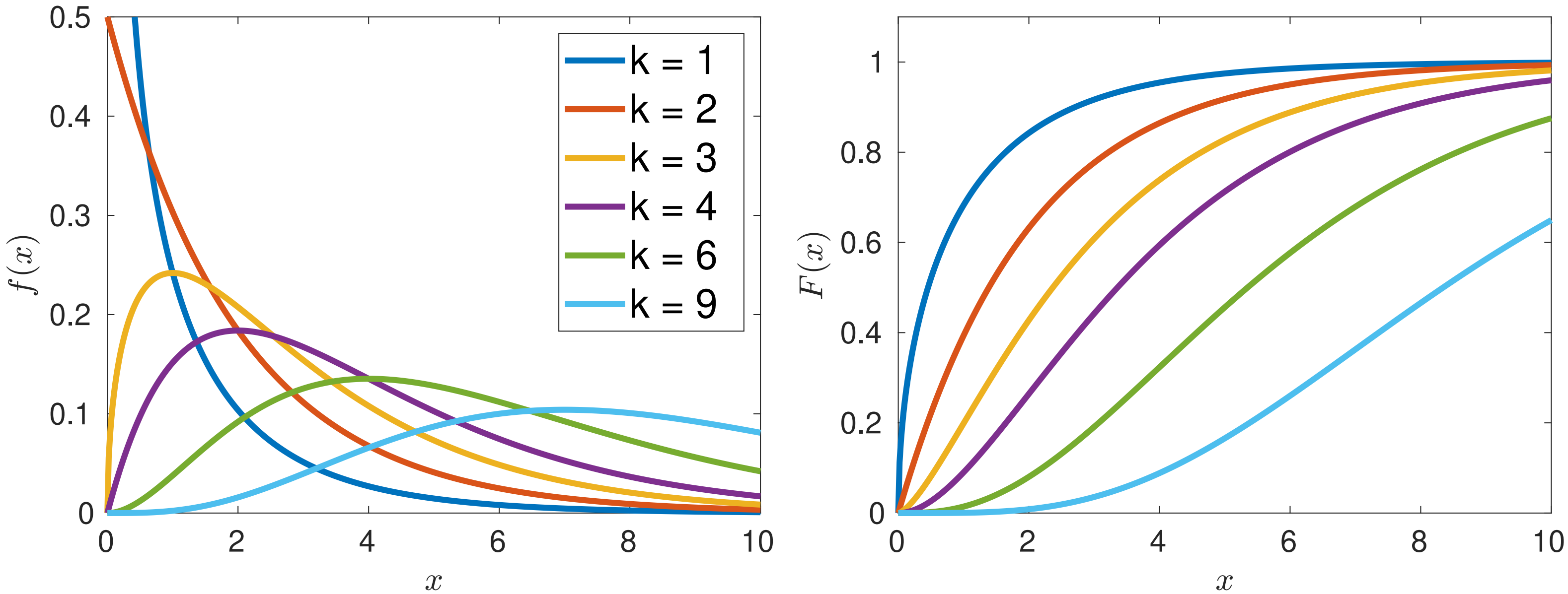
Mean

$$\mathbf{E}[X] = k$$

Variance

$$\mathbf{var}[X] = 2k$$

Example of chi-squared PDF



Rayleigh random variables

- arise when observing the magnitude of a vector
- ex. The absolute values of random complex numbers whose real and imaginary are i.i.d. Gaussian

PDF

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \quad x \geq 0, \quad \alpha > 0$$

Mean

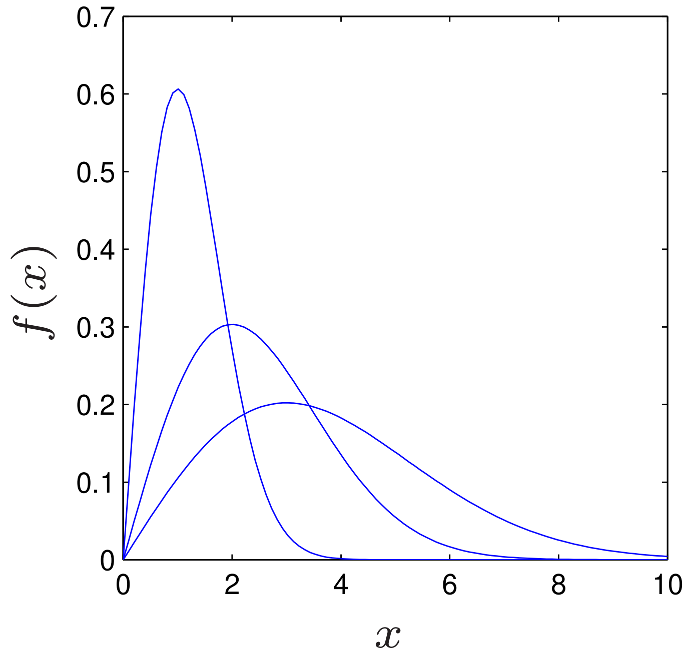
$$\mathbf{E}[X] = \sigma\sqrt{\pi/2}$$

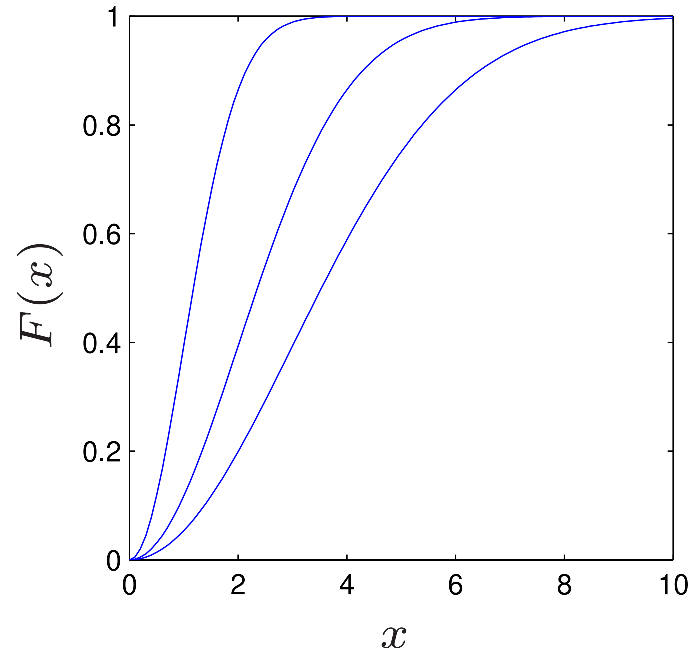
Variance

$$\mathbf{var}[X] = \frac{4-\pi}{2}\sigma^2$$

if X is Rayleigh, then X^2 is $\mathcal{X}_2^2$

Example of Rayleigh PDF





- parameters: $\sigma = 1, 2, 3$

Laplacian random variables

PDF

$$f(x) = \frac{\alpha}{2} e^{-\alpha|x-\mu|}, \quad -\infty < x < \infty$$

Mean

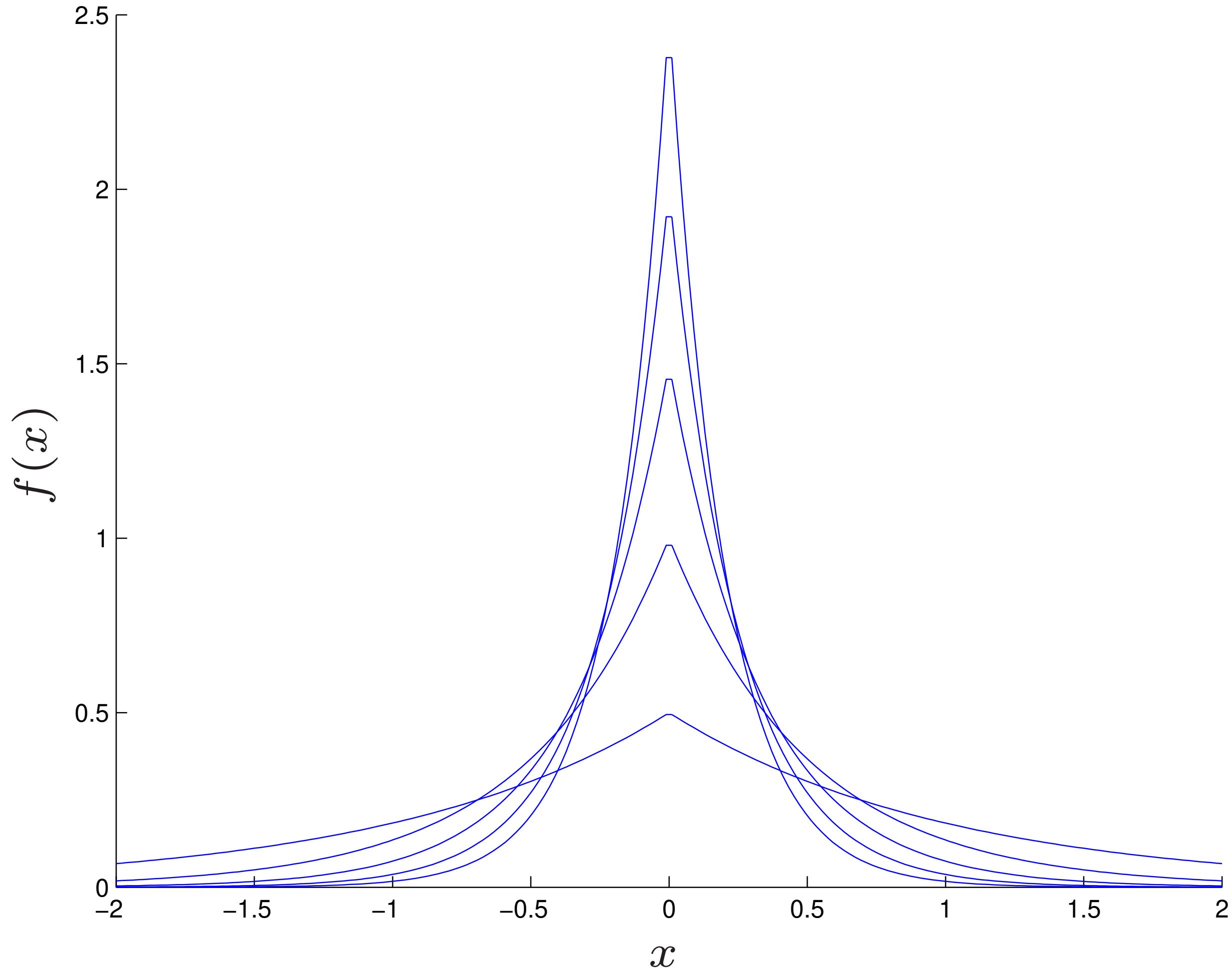
$$\mathbf{E}[X] = \mu$$

Variance

$$\mathbf{var}[X] = \frac{2}{\alpha^2}$$

- arise as the difference between two i.i.d exponential RVs
- unlike Gaussian, the Laplace density is expressed in terms of the *absolute* difference from the mean

Example of Laplacian PDF



- parameters: $\mu = 1$, $\alpha = 1, 2, 3, 4, 5$

Related MATLAB commands

- `cdf` returns the values of a specified cumulative distribution function
- `pdf` returns the values of a specified probability density function
- `randn` generates random numbers from the standard Gaussian distribution
- `rand` generates random numbers from the standard uniform distribution
- `random` generates random numbers drawn from a specified distribution
- `histogram` plots a histogram of data samples

References

Chapter 3,4 in
A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson Prentice Hall, 2009