Probability

Contents

1 Set Theory 2
1.1 Random Variables 2

2 Some Common Distributions 3
2.1 Bernoulli Distribution 3
2.2 Beta Distribution 4
2.3 Binomial Distribution 5
2.4 Bivariate Normal Distribution 6
2.5 Cauchy Distribution 6
2.6 Chi Squared Distribution 7
2.7 Exponential Distribution 8
2.8 F-distribution 9
2.9 Gamma Distribution gamma(α, β) 10
2.10 Geometric Distribution 11
2.11 Normal Distribution n(μ, σ^2) 12
2.12 Poisson Distribution 13
2.13 Uniform 14

1 Set Theory

A **random experiment** is an experiment that produces outcomes which are not predictable with certainty in advance. For instance, tossing a coin 2 times is a random experiment. The **sample space** for a random experiment is the collection of all possible outcomes. We often use the letter S to denote the sample space of a random experiment. For the coin tossing random experiment, we have

$$S = \{\mathrm{HT}, \mathrm{HH}, \mathrm{TT}, \mathrm{TH}\},$$

where H denotes heads and T denotes tails. An **event** is any collection of possible outcomes of an experiment; i.e. a subset of S. We often use the letter A to denote an event.

1.1 Random Variables

Definition 1.1. Let $(S, \mathcal{B}, \mathrm{P})$ be a probability space and let $X\colon S \to \mathbb{R}$ be a random variable. We say

1. X is **continuous** if F_X is continuous
2. X is **discrete** if F_X is a step function
3. X is a **mixture** if F_X contains both continuous and step pieces.
4. We say X has an **exponentional distribution** if

   $$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-x/\beta} & x > 0, \end{cases}$$

   where $\beta > 0$.
5. We say X has a **binomial distribution** if

   $$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

   where n is a positive integer and $0 \leq p \leq 1$. Values such as n and p that can be set to different values, producing different probability distributions, are called **parameters**.
6. If $Y\colon S \to \mathbb{R}$ is another random variable, then we say X and Y are **identically distributed** if

   $$P_X(X \in B) = P_Y(Y \in B)$$

   for all $B \in \mathcal{B}(\mathbb{R})$. It turns out that this is equivalent to

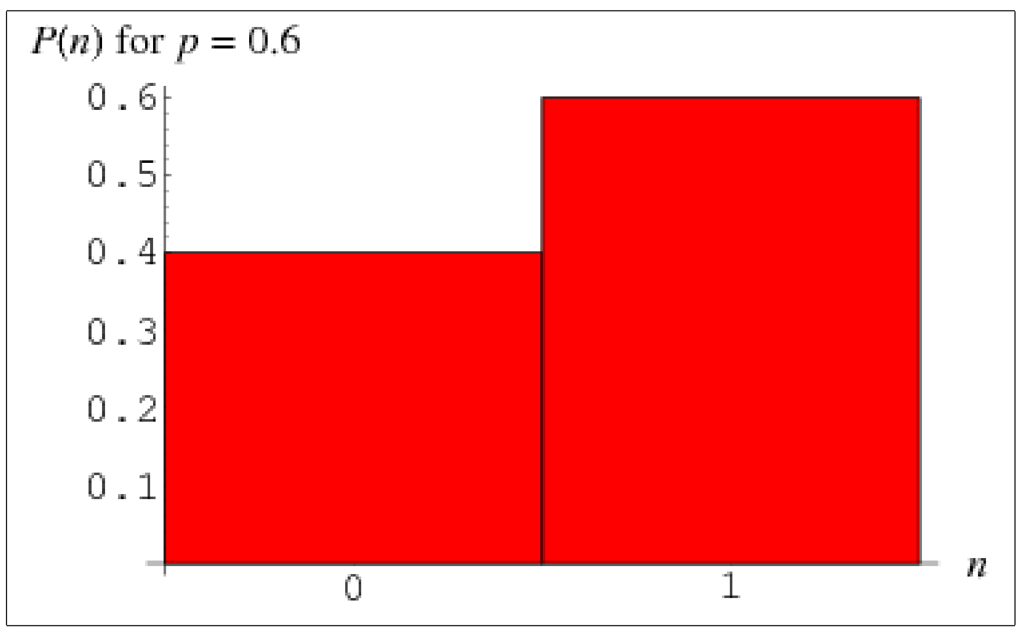
   $$F_X(x) = F_Y(x)$$

   for all $x \in \mathbb{R}$.

2 Some Common Distributions

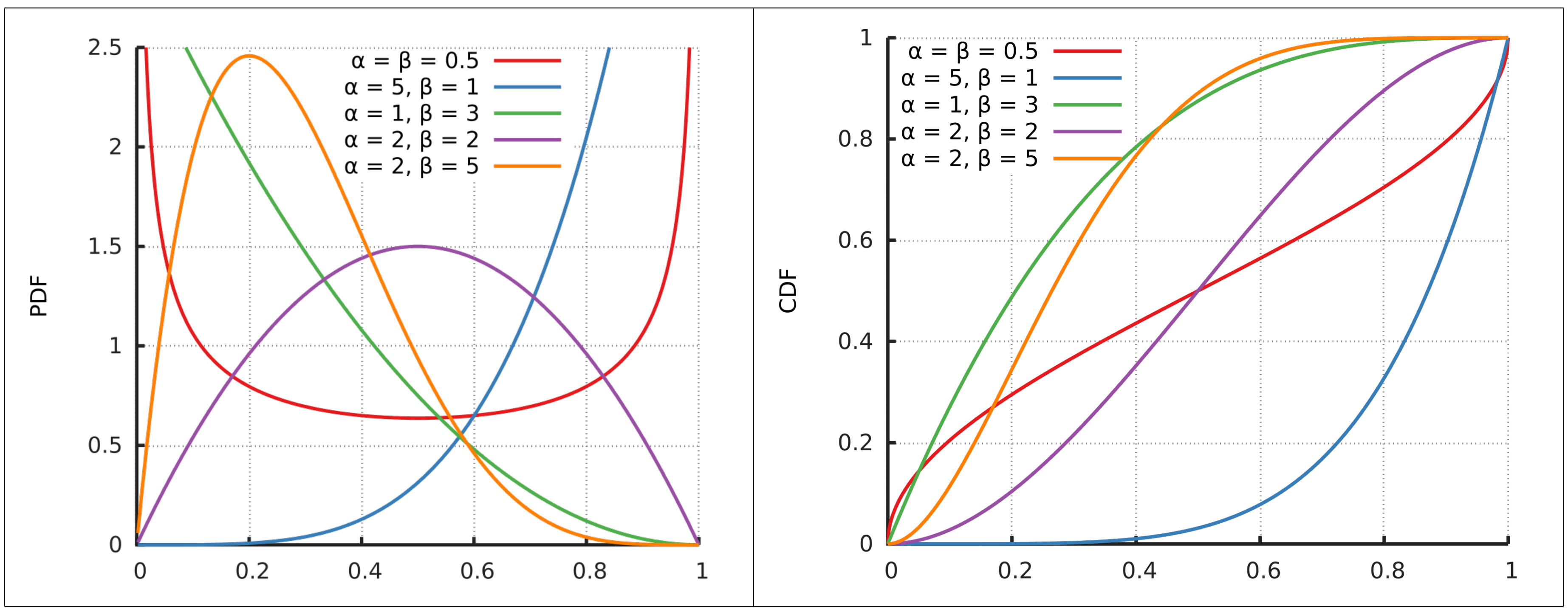
2.1 Bernoulli Distribution

notation	$X \sim \text{Bernoulli}(p)$
pmf/pdf	$f_X(x\|p) = \begin{cases} p^x(1-p)^{1-x} & \text{if } x \in \{0,1\} \text{ (fail/success)} \\ 0 & \text{else} \end{cases}$
parameters	$p \in [0,1]$ "probability of success".
mean	$\text{E}X = p$
variance	$\text{Var}\, X = p(1-p)$
mgf	$M_X(t) = pe^t + (1-p)$



2.2 Beta Distribution

notation	$X \sim \text{beta}(\alpha, \beta)$
pmf/pdf	$f_X(x\|\alpha,\beta) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } x \in [0,1] \text{ where } B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\ 0 & \text{else} \end{cases}$
parameters	$\alpha, \beta \in \mathbb{R}_{>0}$ "shapes of distribution"
mean	$\text{E}X = \frac{\alpha}{\alpha+\beta}$
variance	$\text{Var}\, X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
mgf	$M_X(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}$

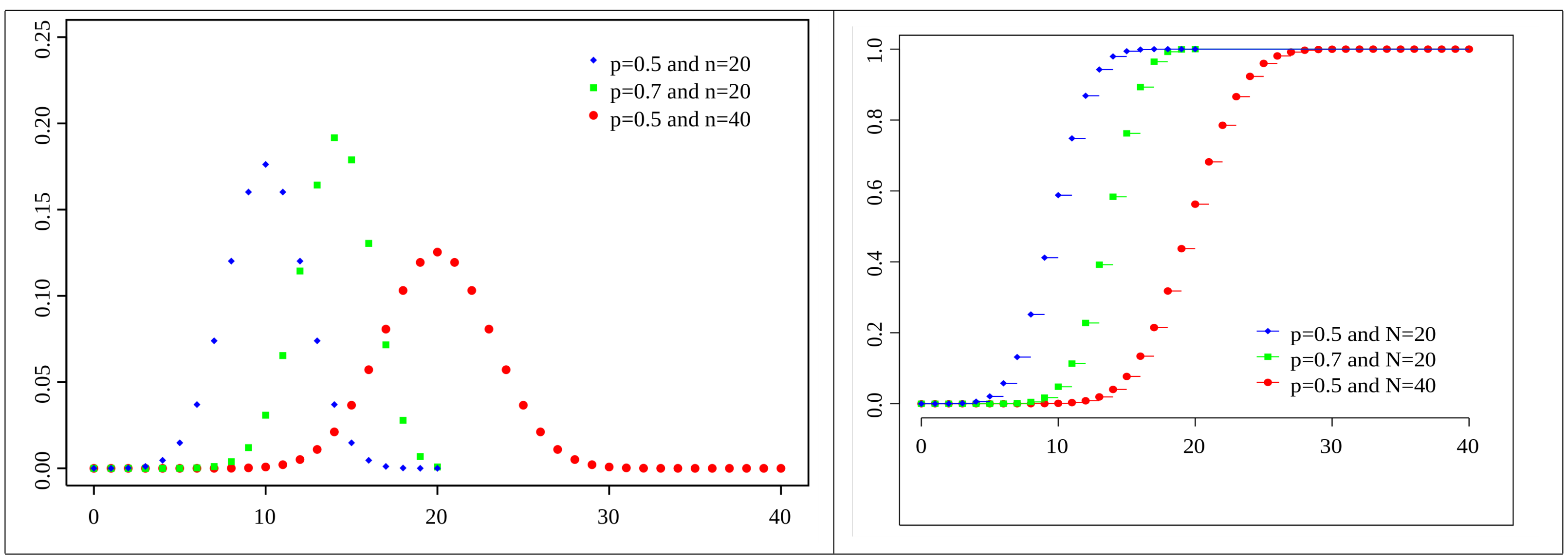


More Properties

- if $X \sim \text{beta}(\alpha, \beta)$, $Y \sim \text{beta}(\alpha+\beta, \gamma)$, and X and Y are independent, then $XY \sim \text{beta}(\alpha, \beta+\gamma)$.

2.3 Binomial Distribution

notation	$X \sim \text{binomial}(n, p)$
pmf/pdf	$f_X(x\|n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x \in \{0, 1, \ldots, n\} \text{ (number of successes)} \\ 0 & \text{else} \end{cases}$
parameters	$n \in \mathbb{Z}_{\geq 0}$ "number of trials" and $p \in [0,1]$ "success probability for each trial".
mean	$\text{E}X = np$
variance	$\text{Var}\, X = np(1-p)$
mgf	$M_X(t) = (pe^t + (1-p))^n$

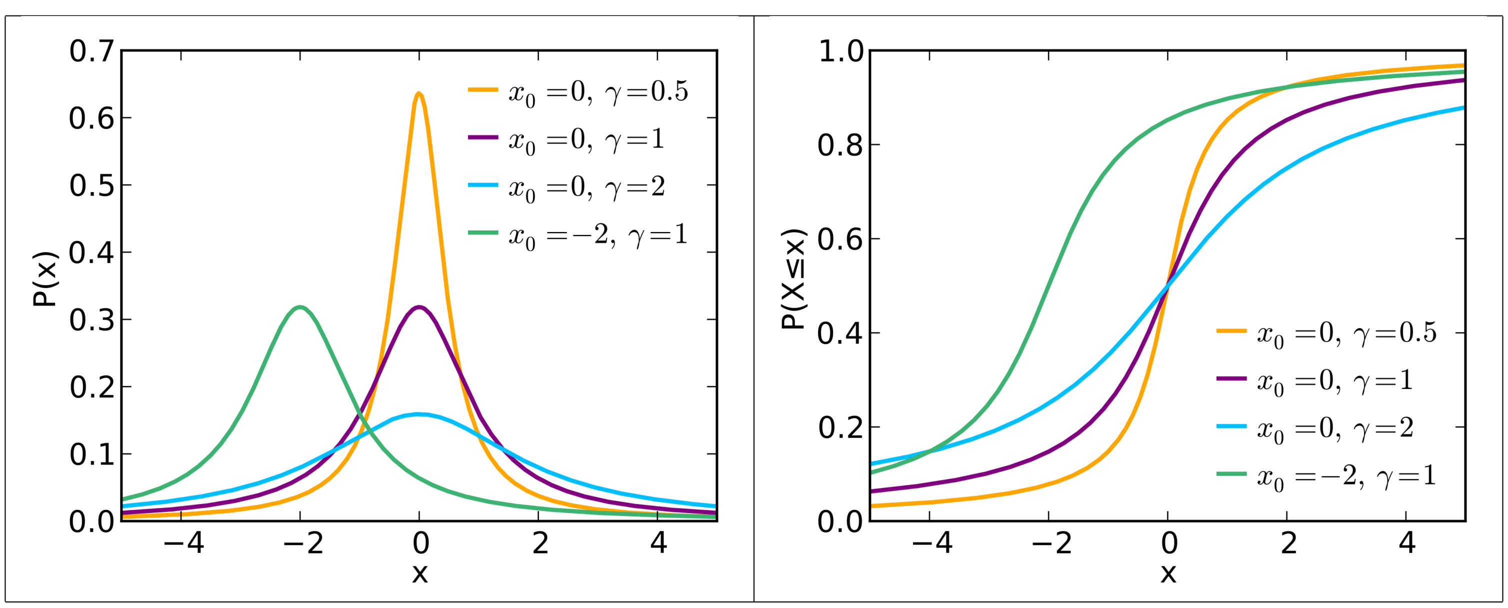


2.4 Bivariate Normal Distribution

notation	$(X, Y) \sim \mathrm{bn}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$
pmf/pdf	$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)}$ for all $(x,y) \in \mathbb{R}^2$
parameters	$\mu_X, \mu_Y \in \mathbb{R}$ (means), $\sigma_X, \sigma_Y \in \mathbb{R}_{>0}$ (square root of variances), and $\rho \in (0,1)$ (correlation)
marginal distribution of X	$\mathrm{n}(\mu_X, \sigma_X^2)$
marginal distribution of Y	$\mathrm{n}(\mu_Y, \sigma_Y^2)$
correlation between X and Y	$\rho_{XY} = \rho$

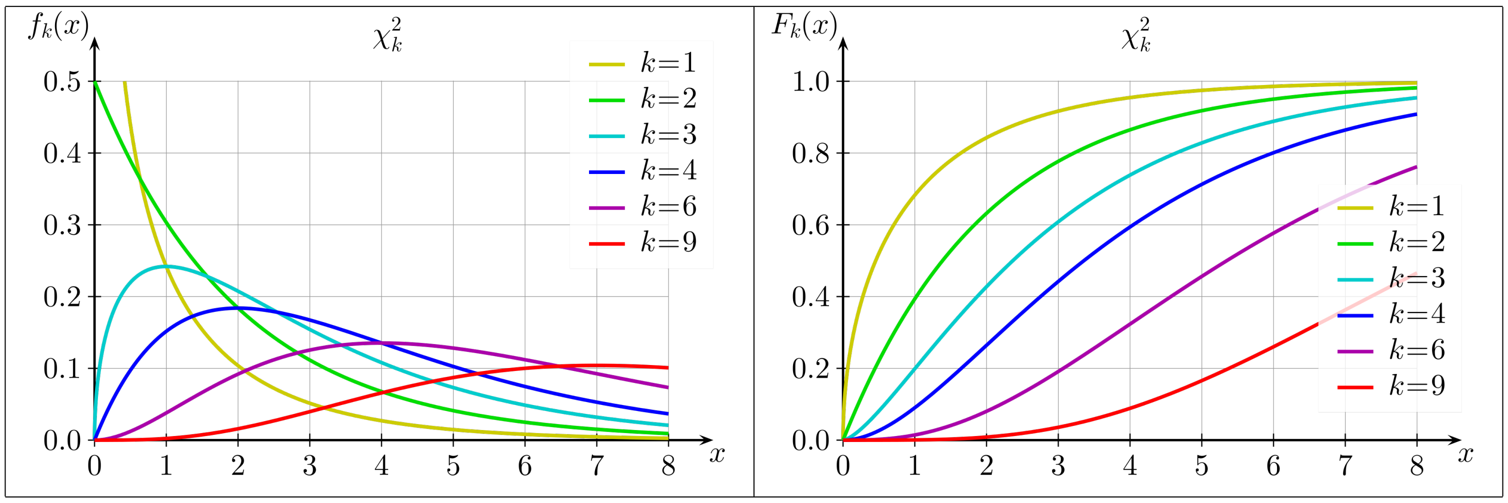
2.5 Cauchy Distribution

notation	$X \sim \mathrm{Cauchy}(\theta, \sigma)$ or $X \sim \mathrm{Cauchy}(x_0, \gamma)$
pmf/pdf	$f_X(x\|\theta,\sigma) = \begin{cases} \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2} & \text{if } x \in \mathbb{R} \\ 0 & \text{else} \end{cases}$
parameters	$\theta \in \mathbb{R}$ location of "center" of pdf and $\sigma \in \mathbb{R}_{>0}$ "spread" of pdf
mean	does not exist
variance	does not exist
mgf	does not exist



2.6 Chi Squared Distribution

notation	$X \sim \chi^2(k)$
pmf/pdf	$f_X(x\|k) = \begin{cases} \frac{1}{\Gamma(k/2)2^{k/2}} x^{(k/2)-1} e^{-x/2} & \text{if } x \in \mathbb{R}_{\geq 0} \\ 0 & \text{else} \end{cases}$
parameters	$k \in \mathbb{Z}_{\geq 1}$ degrees of freedom
mean	$\mathrm{E} X = k$
variance	$\operatorname{Var} X = 2k$
mgf	$M_X(t) = \left(\frac{1}{1-2t}\right)^{k/2}$ where $t < 1/2$



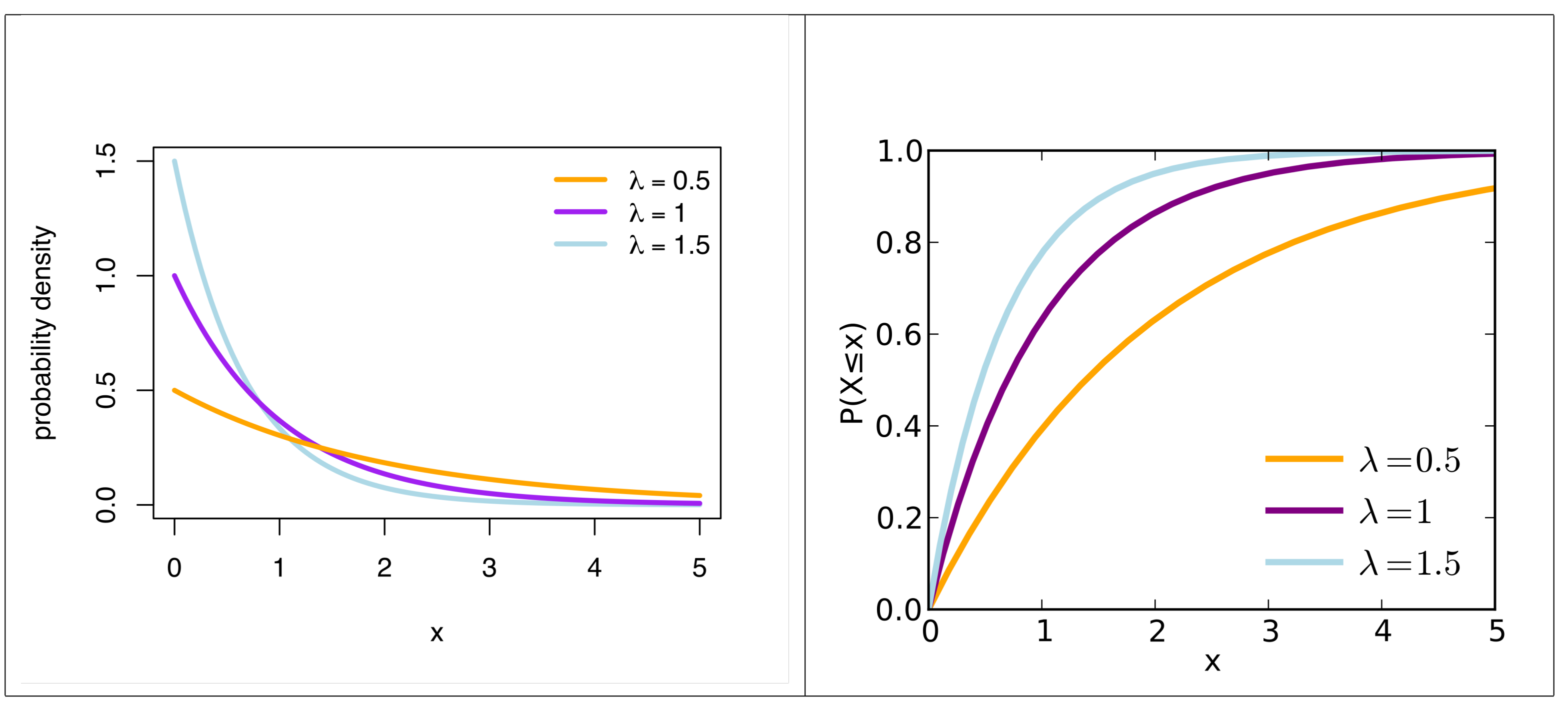
The chi-squared distrubtion with k degrees of freedom is the distrubution of a sum of the squares of k independent standard normal random variables. The chi squared distrubtion is a special case of the gamma distribution and is one of the most widely used probability distrubtions in inferential statistics, notably in hypothesis testing and in construcion of confidence intervals.

2.7 Exponential Distribution

In real-world scenarios, the assumption of a constant rate (or probability per unit time) is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. Similar caveats apply to the following examples which yield approximately exponentially distributed variables:

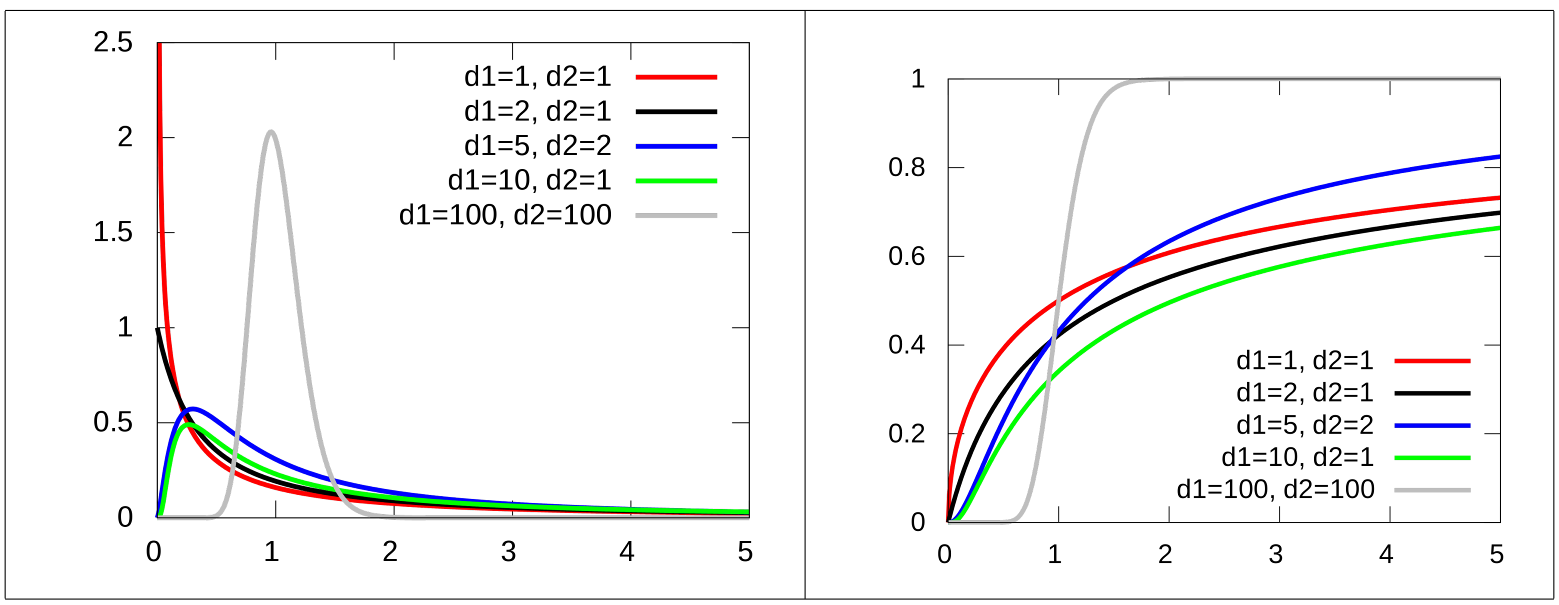
- The time until a radioactive particle decays, or the time between clicks of a Geiger counter.
- The time it takes before your next telephone call.
- The time until default (on payment to company debt holders) in reduced form credit risk modeling

notation	$X \sim \text{exponential}(\beta)$ or $X \sim \text{exponential}(1/\lambda)$ where $\lambda = \beta^{-1}$
pmf/pdf	$f_X(x\|\beta) = \begin{cases} \frac{1}{\beta}e^{-x/\beta} & \text{if } x \in \mathbb{R}_{\geq 0} \\ 0 & \text{else} \end{cases}$
cdf	$F_X(x\|\beta) = \begin{cases} 1 - e^{-x/\beta} & \text{if } x \in \mathbb{R}_{\geq 0} \\ 0 & \text{else} \end{cases}$
parameters	$\beta \in \mathbb{R}_{>0}$ and $\lambda = 1/\beta$ is "expected rate of occurences"
mean	$\mathrm{E}X = \beta$
variance	$\mathrm{Var}\, X = \beta^2$
mgf	$M_X(t) = \frac{1}{1-\beta t}$



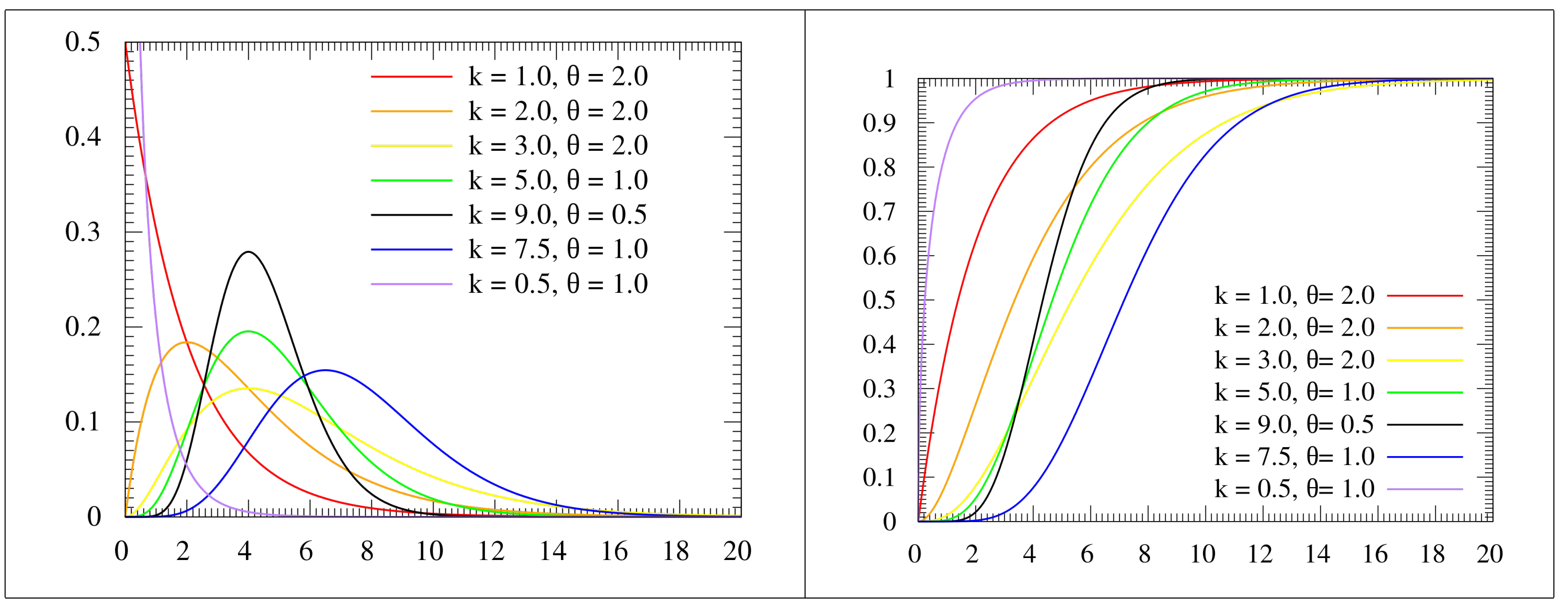
2.8 F-distribution

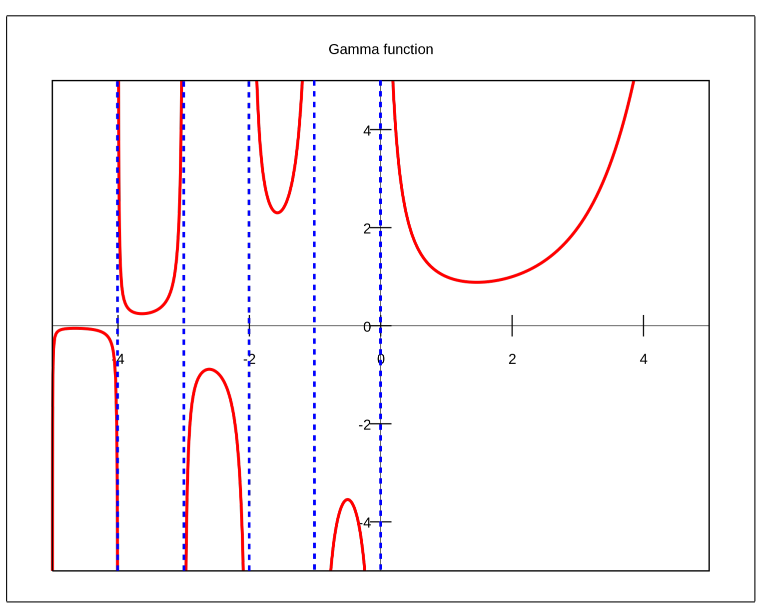
notation	$X \sim \mathrm{F}(d_1, d_2)$
pmf/pdf	$f_X(x\|\alpha,\beta) = \begin{cases} \frac{1}{x\mathrm{B}\left(\frac{d_1}{2},\frac{d_2}{2}\right)}\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}} & x \in \mathbb{R}_{>0} \text{ where } \mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \\ 0 & \text{else} \end{cases}$
cdf	$F_X(x\|\alpha,\beta) = I_{\frac{d_1 x}{d_1 x + d_2}}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)$
parameters	$d_1, d_2 > 0$ degrees of freedom
mean	$\mathrm{E}X = \frac{d_2}{d_2-2}$ for $d_2 > 2$
variance	$\mathrm{Var}\, X = \frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ for $d_2 > 4$
mgf	does not exist



2.9 Gamma Distribution gamma(α, β)

notation	$X \sim \text{gamma}(\alpha, \beta)$ or $X \sim \text{gamma}(k, \theta)$
pmf/pdf	$f_X(x\|\alpha, \beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & x \in \mathbb{R}_{>0} \text{ where } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \mathrm{d}t \\ 0 & \text{else} \end{cases}$
cdf	$F_X(x\|\alpha, \beta) = \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{x}{\beta}\right)$
parameters	$\alpha, \beta \in \mathbb{R}_{>0}$ where α is "shape parameter" and β is "scale parameter"
mean	$\mathrm{E}X = \alpha\beta$
variance	$\mathrm{Var}\, X = \alpha\beta^2$
mgf	$M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha$ (setting $\alpha = 1$ gives us exponential distribution)

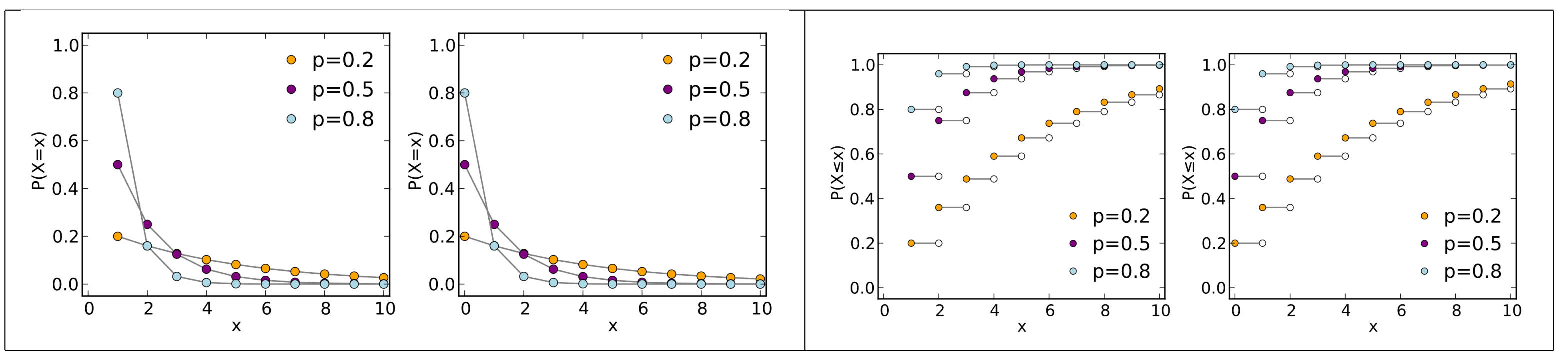




One way to think about this distribution is that it is a "delayed" exponential distrubtion, where a larger value of α corresponds to a longer delay. In particular, if $\alpha = 1$, then this is nothing but the exponential distribution exponential(β), if $\alpha > 1$, then the distrubtion increases a bit before it finally starts to decay, and if $\alpha < 1$, then the distribution is "already decaying" in the sense that it has a pole at $x = 0$.

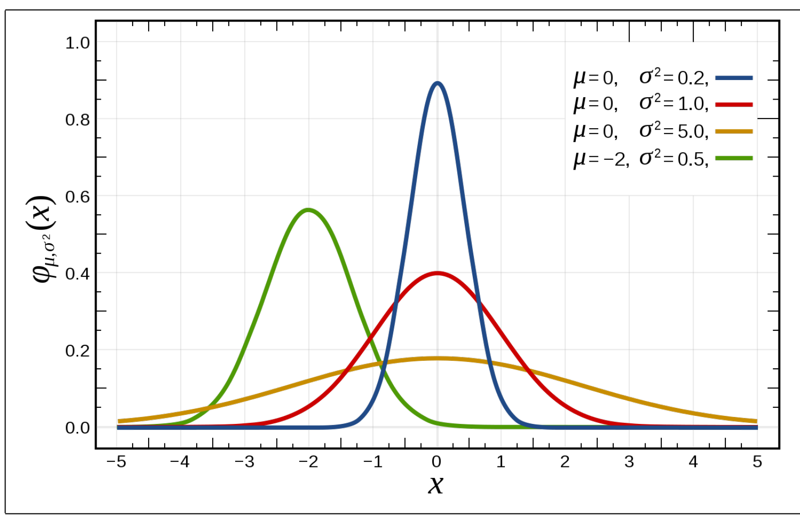
2.10 Geometric Distribution

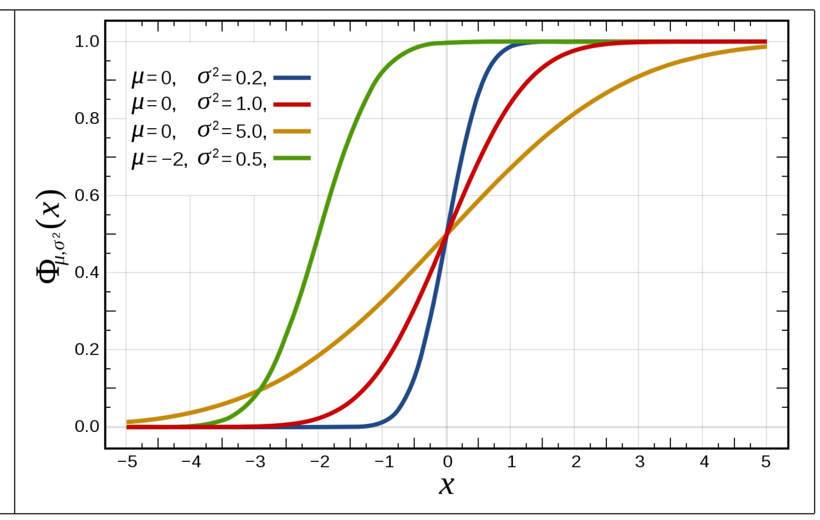
notation	$X \sim \text{geometric}(p)$
pmf/pdf	$\text{P}(X = x \mid p) = \begin{cases} p(1-p)^{x-1} & \text{if } x \in \mathbb{Z}_{\geq 1} \\ 0 & \text{else} \end{cases}$
parameters	$p \in [0,1]$ probability parameter
mean	$\text{E}X = 1/p$
variance	$\text{Var}\, X = (1-p)/p^2$
mgf	$M_X(t) = \frac{pe^t}{1-(1-p)e^t}$ where $t < -\log(1-p)$

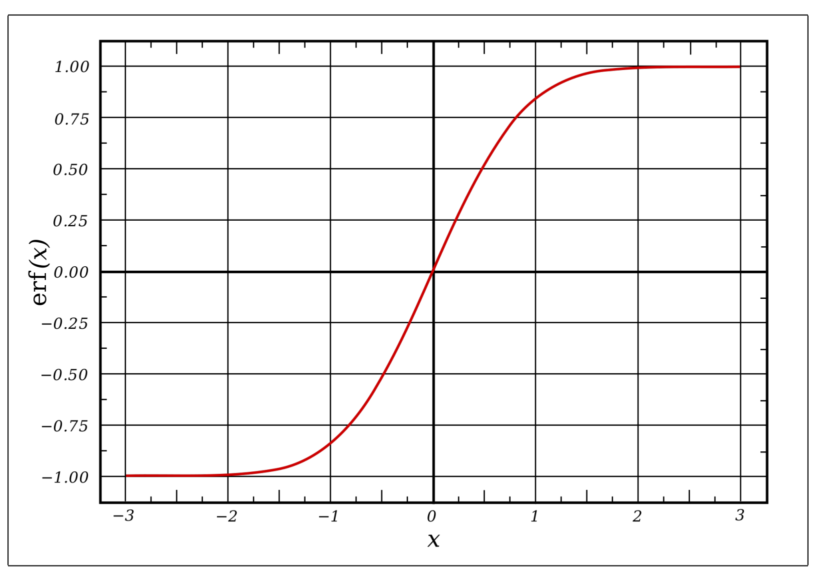


2.11 Normal Distribution $\mathrm{n}(\mu, \sigma^2)$

notation	$X \sim \mathrm{n}(\mu, \sigma^2)$ (we call $\mathrm{n}(0,1)$ the **standard normal** distribution)
pmf/pdf	$f_X(x \mid \mu, \sigma^2) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & \text{if } x \in \mathbb{R} \\ 0 & \text{else} \end{cases}$
cdf	$F_X(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right)$ where $\operatorname{erf} z = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} \mathrm{d}t$.
parameters	$\mu \in \mathbb{R}$ (location) and $\sigma \in \mathbb{R}_{>0}$ (scale)
mean	$\mathrm{E}X = \mu$
variance	$\operatorname{Var} X = \sigma^2$
mgf	$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$





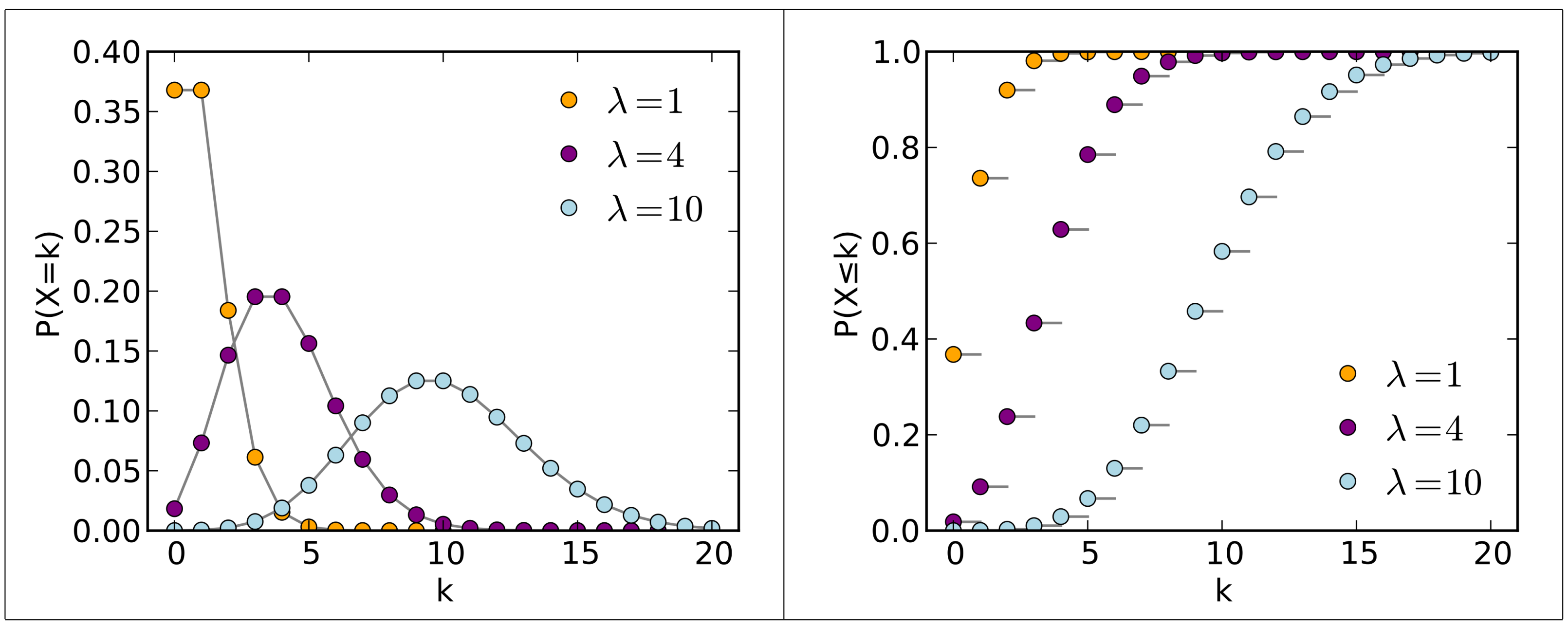


More Properties

- if $X \sim \mathrm{n}(\mu, \sigma^2)$, $Y \sim \mathrm{n}(\gamma, \tau^2)$, and X and Y are independent, then $X + Y \sim \mathrm{n}(\mu + \gamma, \sigma^2 + \tau^2)$.

2.12 Poisson Distribution

notation	$X \sim \text{Poisson}(\lambda)$
pmf/pdf	$f_X(x\|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{if } x \in \mathbb{Z}_{\geq 0} \\ 0 & \text{else} \end{cases}$
cdf	$F_X(t\|\lambda) = \sum_{x=0}^{\lfloor t \rfloor} \frac{e^{-\lambda}\lambda^x}{x!}$
parameters	$\lambda \in \mathbb{R}_{>0}$ "expected rate of occurences"
mean	$\text{E}X = \lambda$
variance	$\text{Var}\, X = \lambda$
mgf	$M_X(t) = e^{\lambda(e^t - 1)}$



More Properties

- if $X \sim \text{Poisson}(\theta)$, $Y \sim \text{Poisson}(\lambda)$, and X and Y are independent, then $X + Y \sim \text{Poisson}(\theta + \lambda)$.

2.13 Uniform

notation	$X \sim \text{uniform}(a, b)$
pmf/pdf	$f_X(x\|a, b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{else} \end{cases}$
parameters	$a, b \in \mathbb{R}$ with $a < b$
mean	$\mathrm{E}X = \frac{a+b}{2}$
variance	$\operatorname{Var} X = \frac{(b-a)^2}{12}$
mgf	$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$