توزیع های آماری

سمینار درس بازشناسی آماری الگو

استاد:
جناب آقای دکتر سید امید شهدی

دانشجو:
حمیدرضا واشقانی فراهانی
۹۲۰۱۵۹۴۱۴

دانشگاه آزاد اسلامی واحد علوم و تحقیقات قزوین
بهار ۱۳۹۳

- در مدل سازی وقایع دنیای واقعی، حالات و شرایطی وجود دارند به گونه ای که می توان عملکرد موجودیت ها درون سیستم را به صورت کامل پیش بینی کرد.
- اما دنیایی که مدل سازی از آن دید به سیستم می نگرد بیشتر مبتنی بر احتمالات است تا قطعیت.
  - برای مدل ساز این تغییرات به صورت اتفاقی رخ می دهد و قابل پیش بینی نیست. با این حال مدل های آماری وجود دارند که بتوان به وسیله آن ها احتمالات را توصیف کرد.
- یک مدل مناسب را می توان با نمونه برداری از رویداد هایی که لزوما باید در مدل حاضر باشند، توسعه داد:

1. برای پارامتر های ورودی توزیع خاصی را انتخاب کرده
2. مدل را با آن پارامتر ها اجرا می شود.
3. از نتایج حاصل از اجرای مدل می توان تشخیص داد تا چه حد توزیع انتخابی بر هدف مورد نظر منطبق بوده است.

متغییر تصادفی گسسته:

اگر X متغیر تصادفی باشد و مقادیر ممکن برای X به صورت محدود و یا قابل شمارش در نظر گرفته شده باشد، گوئیم X یک متغیر تصادفی گسسته است.

مثال: تعداد تقاضا ها برای کاریابی که در هر هفته به بنگاه کاریابی ارائه می شوند را می توان با متغیر تصادفی X توصیف کرد به طوری که:

- R_x = possible values of X (range space of X) = {0,1,2,…}
- p(xi) = probability the random variable is xi = P(X = xi)
- p(xi), i = 1,2, … must satisfy:

1. $p(x_i) \geq 0,\ \text{for all } i$
2. $\sum_{i=1}^{\infty} p(x_i) = 1$

زوج $[x_i, p(x_i)], i = 1,2,\ldots,$ را توزیع احتمال X گویند و $p(x_i)$ را تابع چگالی احتمال گسسته متغیر تصادفی X نامند.

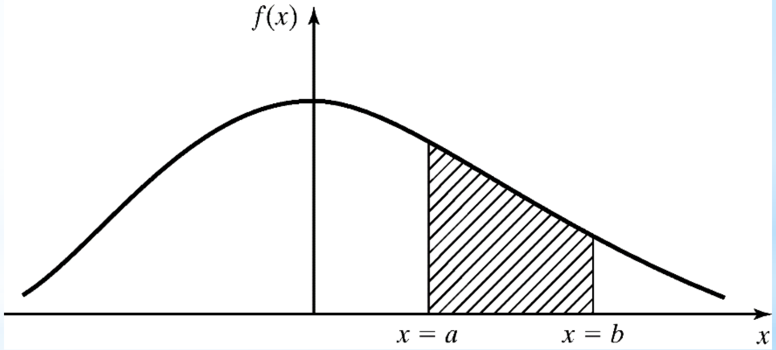
متغیر تصادفی پیوسته

اگر فضای حالت از متغیر تصادفی X به صورت یک بازه یا مجموعه ای از بازه ها باشد گوئیم X یک متغیر تصادفی از نوع پیوسته است.

* احتمال X در بازه [a,b] به صورت: $P(a \le X \le b) = \int_a^b f(x)dx$

* تابع f(x) را تابع چگالی احتمال پیوسته متغیر X نامند. تابع pdf باید شرایط زیر را دارا باشد:

1. $f(x) \ge 0$, for all x in R_X
2. $\int_{R_X} f(x)dx = 1$
3. $f(x) = 0$, if x is not in R_X



- Properties

1. $P(X = x_0) = 0$, because $\int_{x_0}^{x_0} f(x)dx = 0$
2. $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$

تابع توزیع افزایشی

تابع توزیع افزایشی (cdf) را با F(x) نمایش می دهیم و آن را به صورت زیر تعریف میکنیم.

- $F(x) = P(X \le x)$
  - If X is discrete, then $$F(x) = \sum_{\substack{\text{all} \\ x_i \le x}} p(x_i)$$
  - If X is continuous, then $$F(x) = \int_{-\infty}^{x} f(t)dt$$

- Properties

1. F is nondecreasing function. If $a < b$, then $F(a) \le F(b)$
2. $\lim_{x\to\infty} F(x) = 1$
3. $\lim_{x\to-\infty} F(x) = 0$

- تمام مسائل مربوط به توزیع احتمال متغیر تصادفی X را می توان با استفاده از cdf پاسخ داد.

$$P(a < X \le b) = F(b) - F(a), \text{ for all } a < b$$

توزیع های گسسته

1. Bernoulli distribution	1. توزیع برنولی
2. Binomial distribution	2. توزیع دو جمله ای
3. Geometric distribution	3. توزیع هندسی
4. Negative Binomial distribution	4. توزیع دو جمله ای منفی
5. Hypergeometric distribution	5. توزیع فوق هندسی
6. Poisson distribution	6. توزیع پواسون

توزیع های گسسته:

Relative frequency distributions for “counting” experiments.

- Bernoulli Distribution
- Binomial Distribution
- Negative Binomial
- Poisson Distribution
- Geometric Distribution
- Multinomial Distribution

Yes-No responses.

Sums of Bernoulli responses

Number of trials to k^{th} event

Points in given space

Number of trials until first *success*

Multiple possible outcomes for each trial

توزیع برنولی:

یک توزیع گسسته است که مقادیر 1 در صورت موفقیت آزمایش و 0 را در صورت شکست می گیرد.احتمال موفقیت آزمایش برابر p است و احتمال شکست آن برابر $q = 1 - p$ است.

$$P(X=1)=1-P(X=0)=p$$

$$P(X=0)=1-P(X=1)=q$$

$$\Omega=\{S,F\}\Rightarrow P(F)+P(S)=q+p=1$$

آنگاه X یک متغیر تصادفی برنولی با تابع چگالی

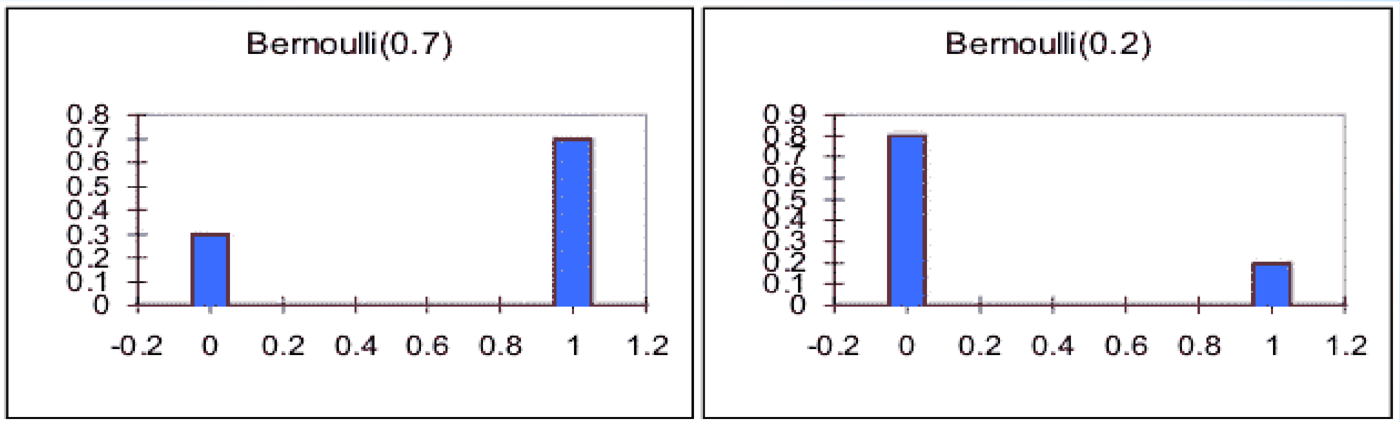
$$P(X=x)=f(x)=\begin{cases} p^x(1-p)^{1-x} & x=0,1 \\ 0 & o.w. \end{cases}$$

می باشدکه $p + q = 1$ است.

$$Var(X)=pq$$

$$E(X)=p$$

$$M_X(t)=(q+pe^t)$$

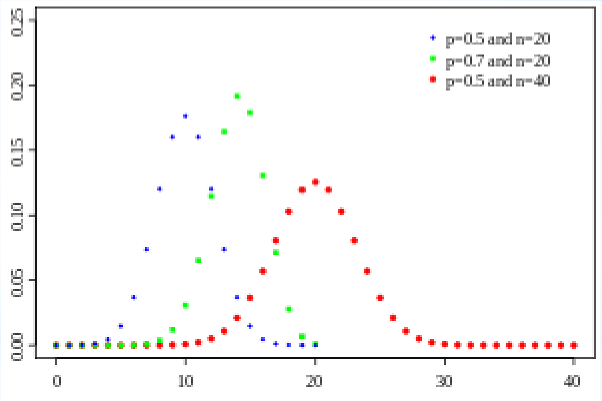


توزیع دو جمله ای:

اگر یک آزمایش برنولی با پارامتر p را n بار مستقلا تکرار کنیم آزمایشی با 2^n برآمد خواهیم داشت،که هر یک دنباله ای n تایی از پیروزی یا شکست می باشد.حال اگر متغیر تصادفی X را تعدادپیروزیهای مشاهده شده در یک دنباله تعریف کنیم، آنگاه Xیک متغیر تصادفی دو جمله ای است با علامت اختصاری $X\sim Bin(n,p)$وتابع چگالی آن به فرم زیر می باشد:

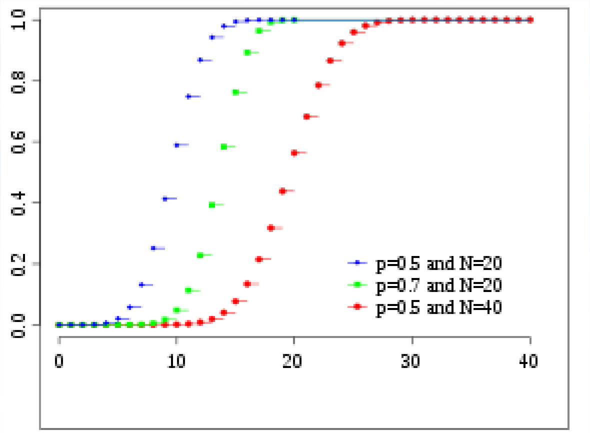
$$f(x;n,p)=P(X=x)=\begin{cases}\binom{n}{x}p^x(1-p^{n-x}) & x=\quad 0,1,2,\dots,n\\ 0 & o.w.\end{cases}$$

$$Var(X)=Var(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}Var(X_i)=npq$$



$$E(X)=E(\sum_{i=1}^{n}X_i)=np$$

$$M_X(t)=(q+pe^t)^n$$



توزیع یکنواخت گسسته :

ساده ترین توزیع احتمال گسسته توزیعی است که در آن متغیر تصادفی X تمام مقادیرش را با احتمالات برابر اختیار کند.اگر یک متغیر تصادفی با تکیه گاه $S_x = \{x_1, x_2, \ldots, x_n\}$ موجود باشد و در تمام نقاط تکیه گاه احتمال ثابت $\frac{1}{N}$ را اختیار کند،آن متغیر تصادفی یکنواخت گسسته می باشد و تابع چگالی(تابع جرم احتمال) آن به فرم زیر است:

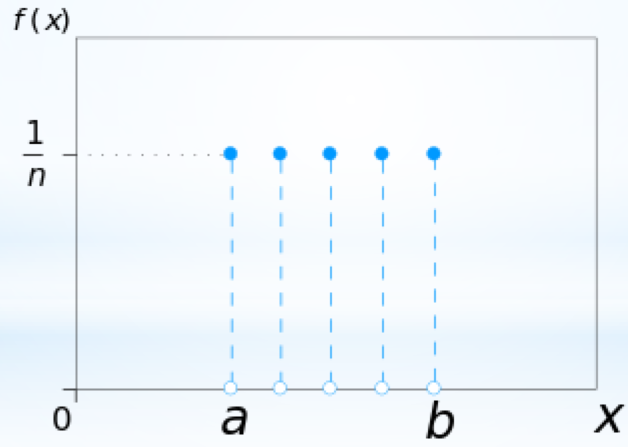
$$f(x) = \begin{cases} \dfrac{1}{N} & x = 1,2,\ldots,N \\ 0 & o.w. \end{cases}$$

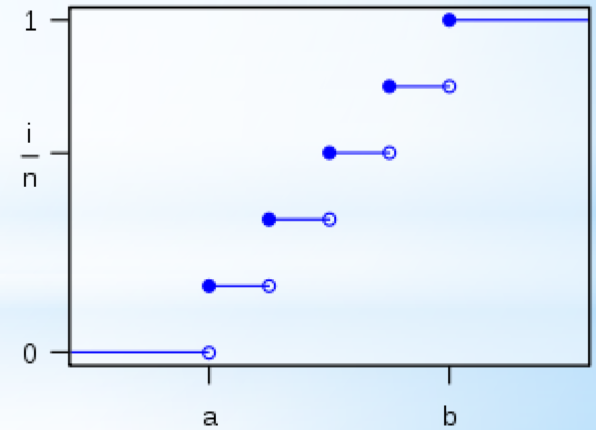
همچنین علامت اختصاری آن به فرم $X \sim DU(N)$ می باشد

$$VAR(X) = \frac{N^2 - 1}{12}$$

$$E(X) = \frac{N+1}{2}$$

$$M_X(t) = E(e^{tx}) = \frac{e^t(1-e^{Nt})}{N(1-e^t)}$$





توزیع هندسی:

اگر یک آزمایش برنولی با شانس p را آنقدر تکرار کنیم تا برای اولین بار پیروزی مشاهده شود، آنگاه یک آزمایش هندسی داریم .

اگر X را تعداد باختها تا مشاهده اولین پیروزی و Y تعداد آزمایشهای لازم تا رسیدن به اولین پیروزی در نظر بگیریم ،آنگاه X توزیع هندسی نوع اول که تابع چگالی آنها به صورت زیر می باشد:

$$f(x) = \begin{cases} pq^x & x = 0,1,2,\ldots \\ 0 & o.w. \end{cases}, p + q = 1$$

و Y توزیع هندسی نوع دوم دارد که تابع چگالی آن به صورت زیر است:

$$f(y) = \begin{cases} pq^{y-1} & y =,1,2,\ldots \\ 0 & o.w. \end{cases}, p + q = 1$$

و با در نظر گرفتن $Y = 1 + X$ می توان هندسی نوع اول را به هندسی نوع دوم تبدیل کرد.

اغلب به متغیر تصادفی هندسی ،متغیر تصادفی زمان انتظار گسسته اطلاق می شود.

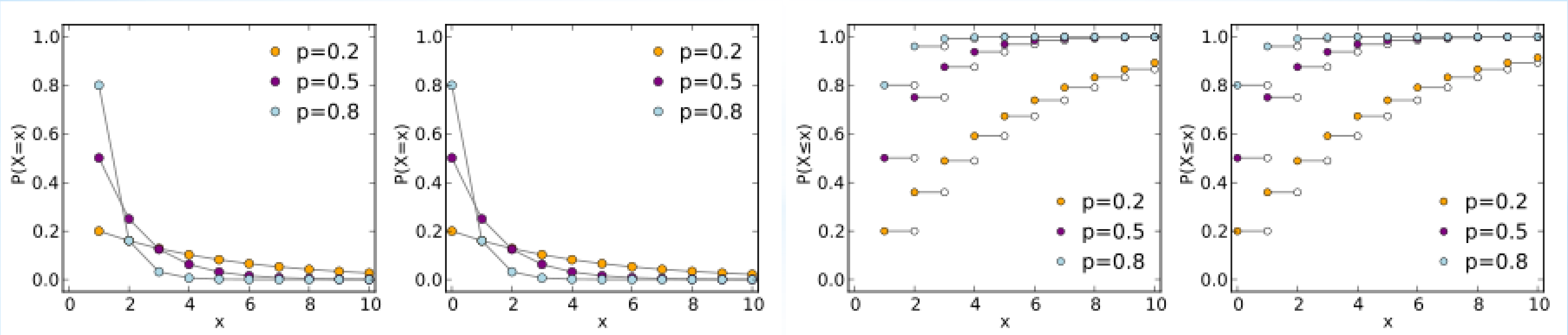
اگر $X\sim Ge(\mathrm{p})$ باشد :

آنگاه برای هندسی نوع اول داریم:

$$E(X)=\frac{q}{p} \quad ,Var(x)=\frac{q}{p^2} \quad ,M_X(t)=E(e^{tX})=\sum_{x=0}^{+\infty}e^{tx}pq^x=p\sum_{x=0}^{+\infty}(e^tq)^x=\frac{p}{1-qe^t}$$

و برای هندسی نوع دوم

$$E(Y)=\frac{1}{p} \quad ,Var(Y)=\frac{q}{p^2} \quad ,M_Y(t)=M_{X+1}(t)=e^tM_X(t)=\frac{pe^t}{1-qe^t}$$



توزیع دو جمله ای منفی:

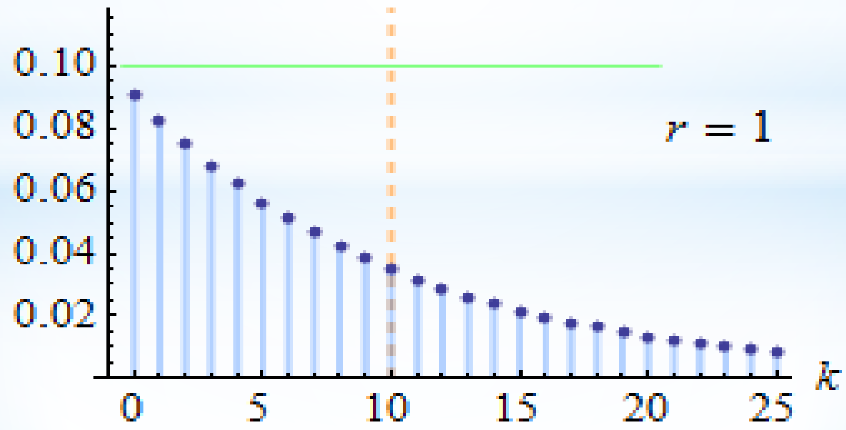
اگر یک آزمایش برنولی با شانس P را آنقدر تکرار کنیم تا به r موفقیت برسیم یک آزمایش تصادفی انجام داده ایم که در آن X تعداد باختهای مشاهده شده تا مشاهده r پیروزی وY تعداد آزمایشها لازم تا مشاهده r پیروزی می باشد.آنگاه X و Y توزیع دو جمله ای منفی نوع اول و نوع دوم دارند.که تابع چگالی آن بدین صورت   می باشد. $X \sim NB(r,p)$

دوجمله ای منفی نوع اول :

$$f(x)=\begin{cases}\binom{r+x-1}{x}p^r q^x & x=0,1,\dots \\ 0 & o.w.\end{cases}, p+q=1 \qquad E(X)=\frac{rq}{p}, Var(X)=\frac{rq}{p^2}, M_X(t)=\left[\frac{p}{1-qe^t}\right]^r$$

دو جمله ای منفی نوع دوم :

$$f(y)=\begin{cases}\binom{y-1}{r-1}p^r q^{y-r} & y=r,r+1,\dots \\ 0 & o.w.\end{cases}, p+q=1 \qquad E(X)=\frac{r}{p}, Var(X)=\frac{rq}{p^2}, M_X(t)=\left[\frac{pe^t}{1-qe^t}\right]^r$$



توزیع فوق هندسی:

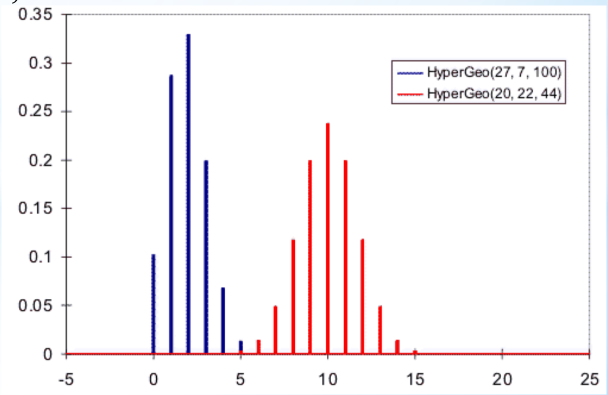
آزمایشی که از انجام چند آزمایش غیر مستقل برنولی بدست می آید آزمایش فوق هندسی نام دارد.به عبارت دیگر فرض کنید آزمایشات برنولی تکرار شده مستقل از هم نباشد آنگاه آزمایش به دست آمده دیگر دو جمله ای نیست .

به طور کلی اگر جمعیتی شامل N عضو باشد و M عضو آن دارای یک ویژگی خاص باشند و نمونه ای به حجم n بدون جایگذاری از جامعه انتخاب کنیم . آزمایش ما فوق هندسی می باشد.

اگر متغیر تصادفی X را تعداد موفقیتها (تعداد عضوهایی که با ویژگی M در نمونه ظاهر می شوند) متغیر تصادفی فوق هندسی می باشد .

با علامت اختصاری $X\sim HG(N,M,n)$ نشان داده می شود .و تابع چگالی آن عبارت است از :

$$P(X=x)=\begin{cases} \dfrac{\binom{M}{x}\times\binom{N-M}{n-x}}{\binom{N}{n}} & x=\max\{0,n-(N-M)\},\ldots\ldots,\min\{n,M\} \\ 0 & o.w. \end{cases}$$



$$Var(X)=n\times\frac{M}{N}\times\frac{N-M}{N}\times\frac{N-n}{N-1} \qquad E(X)=n\times\frac{M}{N}$$

توزیع پواسون:

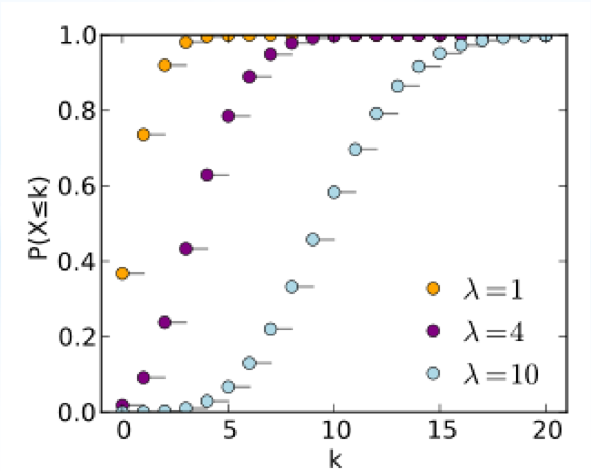
آزمایشی که تعداد موفقیتها را در یک فاصله زمانی و یا مکانی می شمارد و در اصول سه گانه زیر صدق می کند آزمایش پواسن نام دارد.:

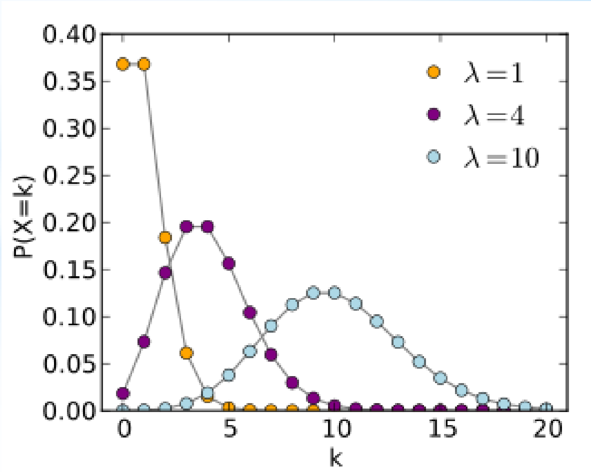
1. تعداد موفقیتها در یک فاصله زمانی کوچک مانند h مستقل از تعداد موفقیتها در فاصله زمانی دیگر است . (اصل استقلال)
2. احتمال اینکه دقیقا" یک واقعه در فاصله زمانی کوتاه به طول hرخ دهد .تقریبا معادل با $\mathrm{h} \times \nu$باشد. (اصل ایستایی)
3. احتمال رخ داد بیش از یک واقعه در فاصله زمانی کوتاه h قابل صرفنظر کردن باشد.(اصل کم یابی)

به νمیانگین نرخ توزیع می گویند

از توزیع پواسن به خاطر ماهیتش می توان جهت شمارش حوادث استفاده کرد.

$$f(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & x = 0,1,2,\ldots \\ 0 & \end{cases}$$





توزیع های پیوسته

1. Uniform distribution	1. توزیع یکنواخت
2. Triangular distribution	2. توزیع مثلثی
3. Normal distribution	3. توزیع نرمال
4. Cauchy distribution	4. توزیع کوشی
5. Exponential distribution	5. توزیع نمایی
6. Lognormal distribution	6. توزیع Lognormal
7. Gamma distribution	7. توزیع گاما
8. Beta distribution	8. توزیع بتا
9. Weibull distribution	9. توزیع وایبول

توزیع های پیوسته:

- **Normal Distribution**
- Log Normal Distribution
- Gamma Distribution
- **Chi Square Distribution**
- **F Distribution**
- **t Distribution**
- Weibull Distribution
- Extreme Value Distribution (Type I and II)

Foundations for much of statistical inference

Environmental variables

Time to failure, radioactivity

Basis for statistical tests.

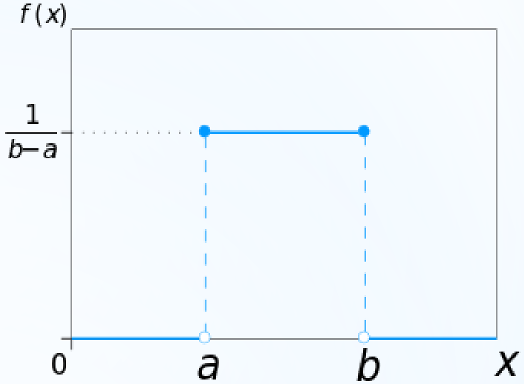
Lifetime distributions

Continuous random variables are defined for continuous numbers on the real line. Probabilities have to be computed for all possible sets of numbers.

توزیع یکنواخت پیوسته :

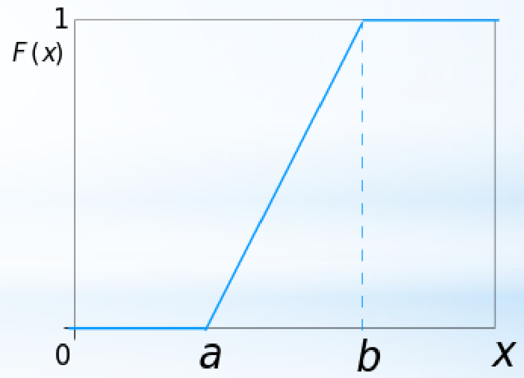
متغیر تصادفی دارای توزیع یکنواخت پیوسته درفاصله (a,b) است و آنرا با $X\sim U(a,b)$ نشان می دهند . هرگاه تابع چگالی احتمال(Probability density function) آن به فرم زیر باشد

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & o.w.\end{cases}$$



همچنین تابع توزیع افزایشی (Cumulative distribution function) آن نیز به فرم زیر می باشد

$$F(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x<b\\ 1 & x\ge b\end{cases}$$



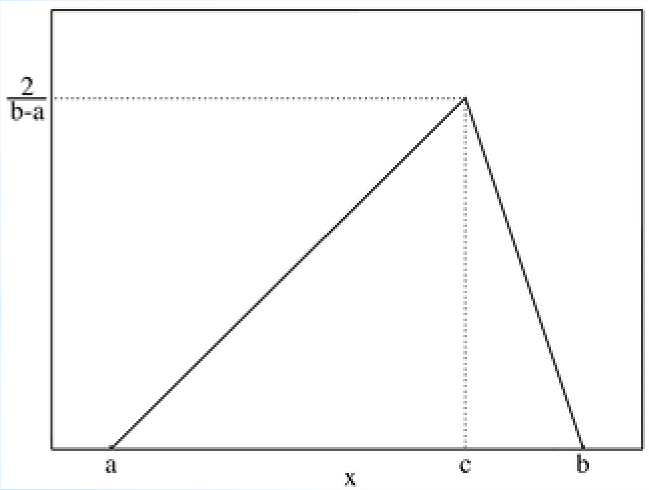
$$E(X)=\frac{a+b}{2}\quad ,Var(X)=\frac{(a-b)^2}{12}\quad ,M_X(t)=\frac{e^{bt}-e^{at}}{(b-a)t}$$

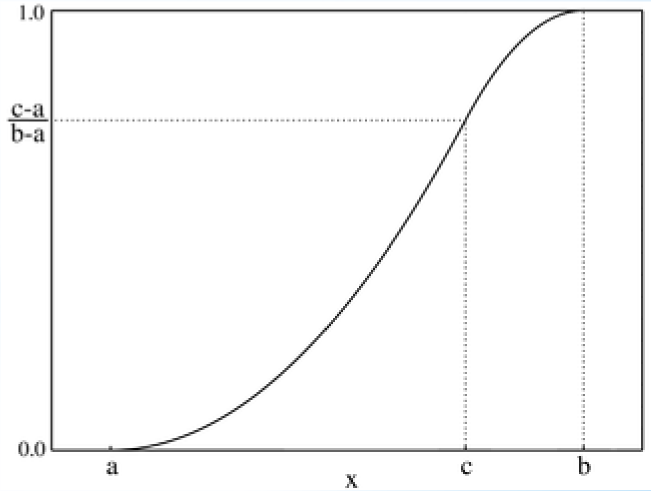
توزیع مثلثی:

اگر X یک متغیر پیوسته با تابع چگالی زیر باشد:

$$f(x;a,b,c)=\begin{cases} 0 & for\ x<a \\ \dfrac{2(x-a)}{(b-a)(c-a)} & for\ a\le x\le b \\ \dfrac{2(b-x)}{(b-a)(b-c)} & for\ c<x<b \\ 0 & for\ b<x \end{cases}$$

این توزیع به مقادیر a و b محدود بوده و مد آن معادل c می باشد بطوریکه $a<b$و $a\le c\le b$ می باشد.





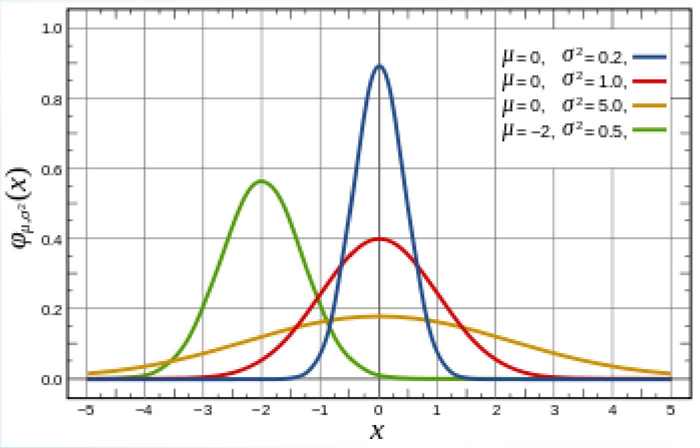
توزیع نرمال:

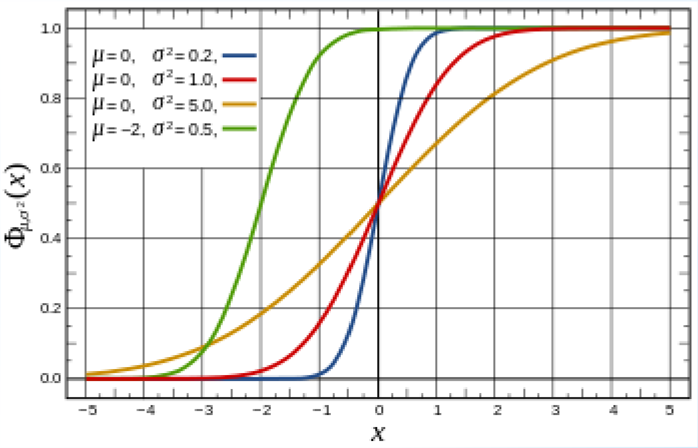
اگر متغیر تصادفی دارای تابع چگالی زیر باشد:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in R, \mu \in R, \sigma > 0$$

در این صورت X توزیع نرمال دارد . آن را با علامت اختصاری $X \sim N(\mu,\sigma^2)$ نشان می دهند.

μ میانگین (پارامتر مکان) و σ انحراف معیار (پارامتر مقیاس) است.





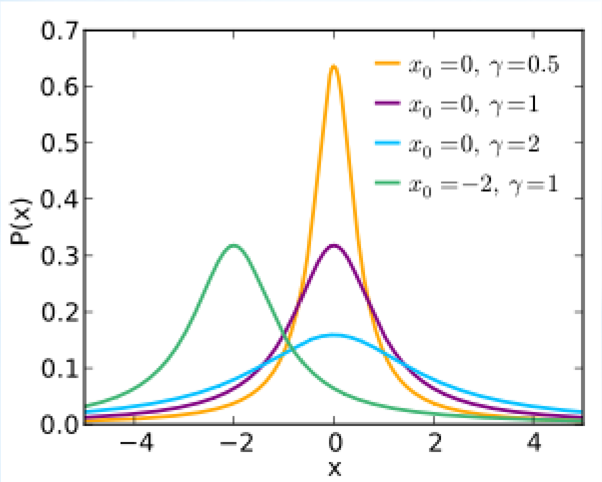
توزیع کوشی:

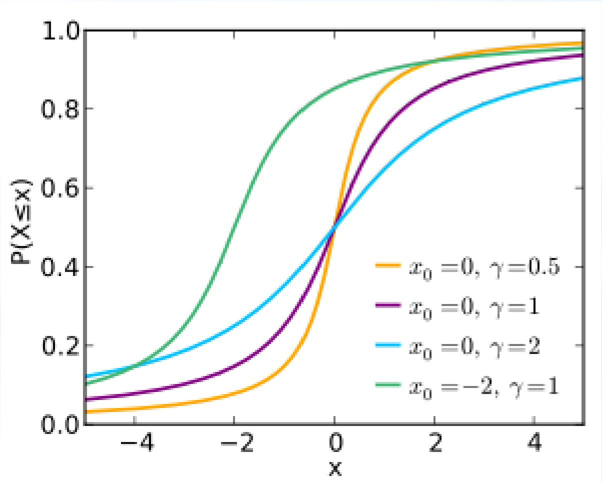
اگر X یک متغیر پیوسته باشد با تابع چگالی زیر باشد:

$$f_X(x;\alpha,\beta)=\frac{1}{\pi\beta\left[1+\left(\frac{x-\alpha}{\beta}\right)^2\right]}\qquad \begin{array}{l} x,\alpha\in R \\ \beta>0 \end{array}$$

دراین صورت X دارای توزیع کوشی با پارامترهای α,β است و آنرا با نماد $X\sim C(\alpha,\beta)$ نشان می دهیم.

نمودار تابع چگالی کوشی متقارن است اما تفاوت آن با نرمال استاندارد این است که احتمال در دمهای منحنی بیشتر ودروسط کمتر است .





توزیع نمایی:

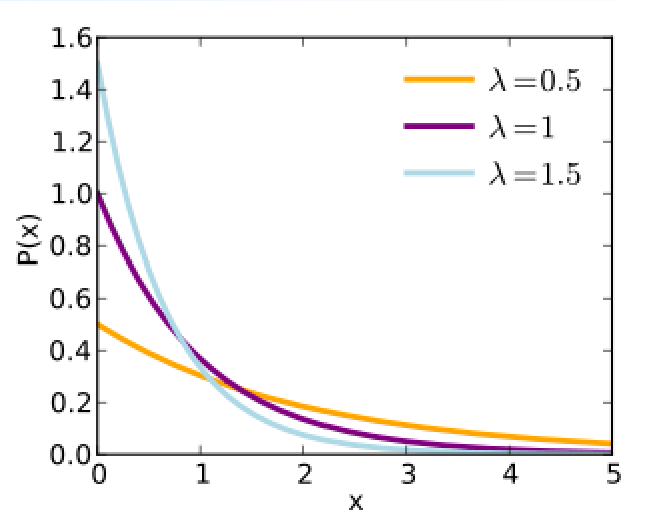
اگر متغیر تصادفی Xدارای تابع چگالی زیر باشد:

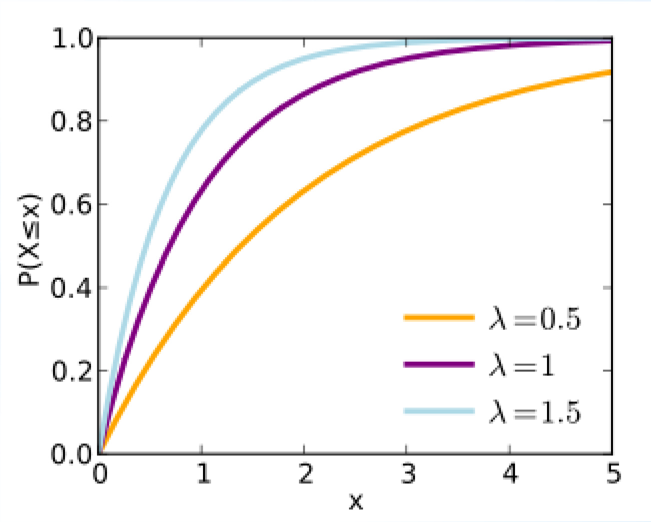
$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & o.w.\end{cases}$$

آنگاه X یک متغیر تصادفی نمایی با پارامتر λاست و آن را با علامت اختصاری $X\sim Exp(\lambda)$نشان می دهند.همچنین تابع توزیع آن عبارت است از:

$$F(x)=\begin{cases}1-e^{-\lambda x} & x\geq 0\\ 0 & o.w.\end{cases}$$

$$E(X)=\frac{1}{\lambda},Var(X)=\frac{1}{\lambda^2},M_X(t)=\frac{\lambda}{\lambda-t}$$





توزیع lognormal:

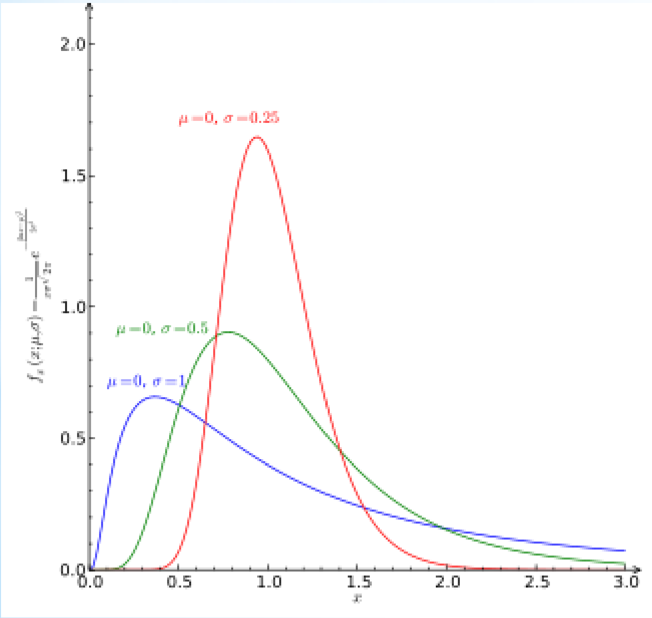
در آمار و احتمال، توزیعی است که لگاریتم طبیعی آن دارای توزیع نرمال با پارامترهای μو σمی‌باشد.

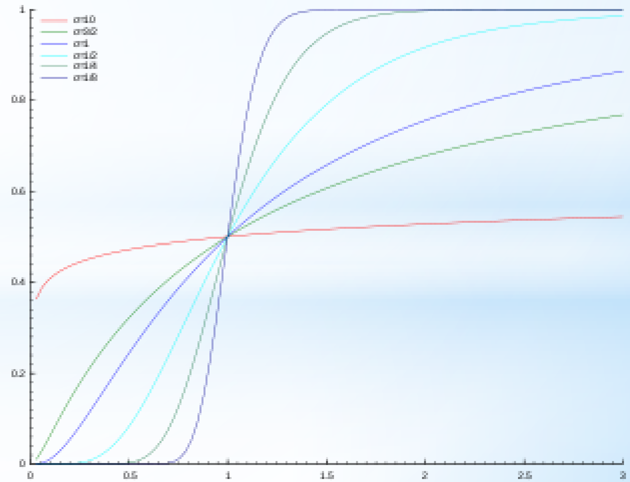
به عبارت دیگر اگر X متغیری با توزیع نرمال باشد، آنگاه $Y = \exp(X)$دارای توزیع لگ نرمال است.

توزیع لگ نرمال برای مدل سازی متغیرهای طبیعی استفاده می شود که از تعدادی از متغیرهای طبیعی دیگر تولید می شوند.

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(\ln[x]-\mu_1)^2}{2\sigma_1^2}\right]$$

$$\mu_1 = \ln\left[\frac{\mu^2}{\sqrt{\sigma^2+\mu^2}}\right], \sigma_1 = \sqrt{\ln\left[\frac{\sigma^2+\mu^2}{\mu^2}\right]}$$





توزیع گاما:

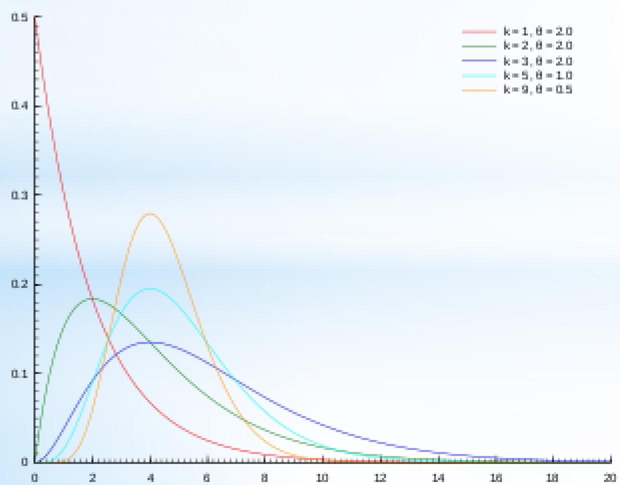
اگرمتغیرتصادفی پیوسته X دارای تابع چگالی زیر باشد:

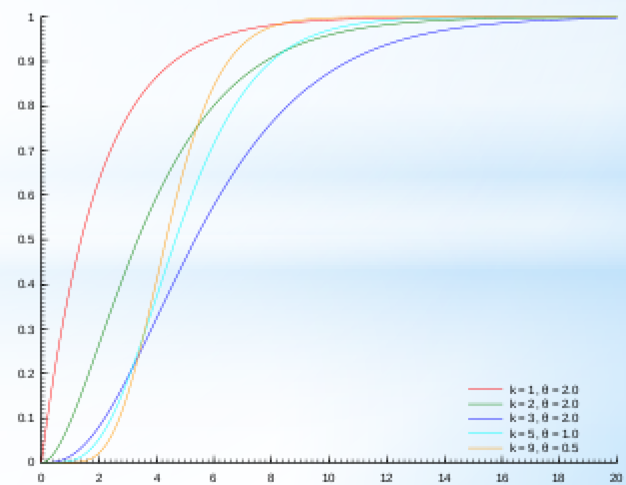
$$f_X(x)=\begin{cases}\dfrac{\lambda^n}{\Gamma(n)}x^{n-1}e^{-\lambda x} & x,\lambda>0\\ 0 & o.w.\end{cases}$$

$$Var(X)=n\theta^2 \qquad E(X)=n\theta \qquad M_X(t)=(1-\theta t)^{-n},t<\frac{1}{\theta}$$

در این صورت می گوییم X دارای توزیع گاما با پارامترهایλ و n بوده و آنرا با نماد$X{\sim}Gamma(n,\lambda)$ نشان میدهیم . علامت Γ مربوط به تابع گاما در ریاضی می باشدکه بصورت زیر تعریف می گردد.

$$\Gamma(n)=\int_0^{+\infty}t^{n-1}e^{-t}dt$$





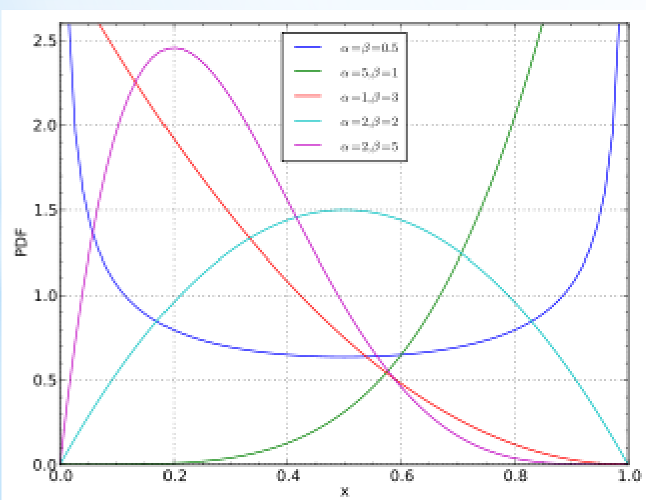
توزیع بتا:

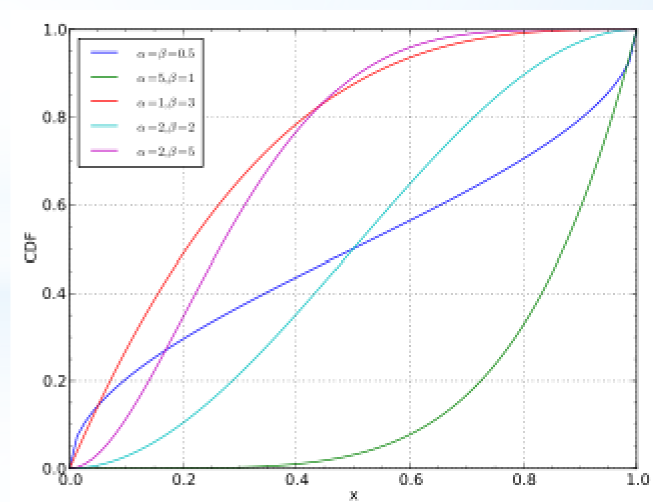
یکی از توزیع های که بیشتر در آمار ناپارامتری (آمار توزیع آزاد)بکار می رود ، توزیع بتا می باشد.اگر متغیر پیوسته X دارای تابع چگالی زیر باشد:

$$f(x)=\begin{cases}\dfrac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1} & 0<x<1 \quad a,b>0\\ 0 & o.w.\end{cases}$$

بطوریکه $\beta(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ در اینصورت می گوییم X دارای توزیع بتا با پارامترهای (a,b) بوده و آنرا با نماد $X\sim\beta(a,b)$ نمایش میدهند.

تابع بتا در ریاضی با پارامترهای aو b به صورت $\beta(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}dt$ تعریف می شود.





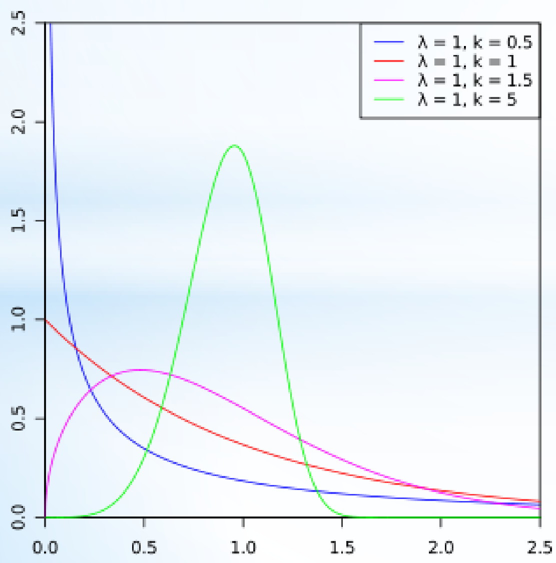
توزیع وایبول:

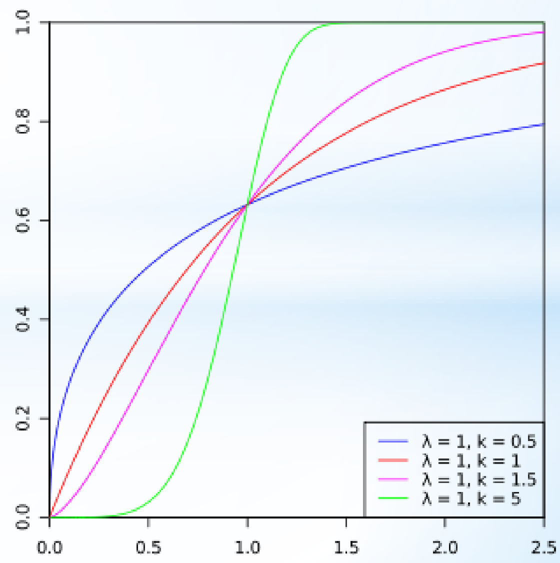
اگرمتغیرتصادفی پیوسته X دارای تابع چگالی زیر باشد:

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

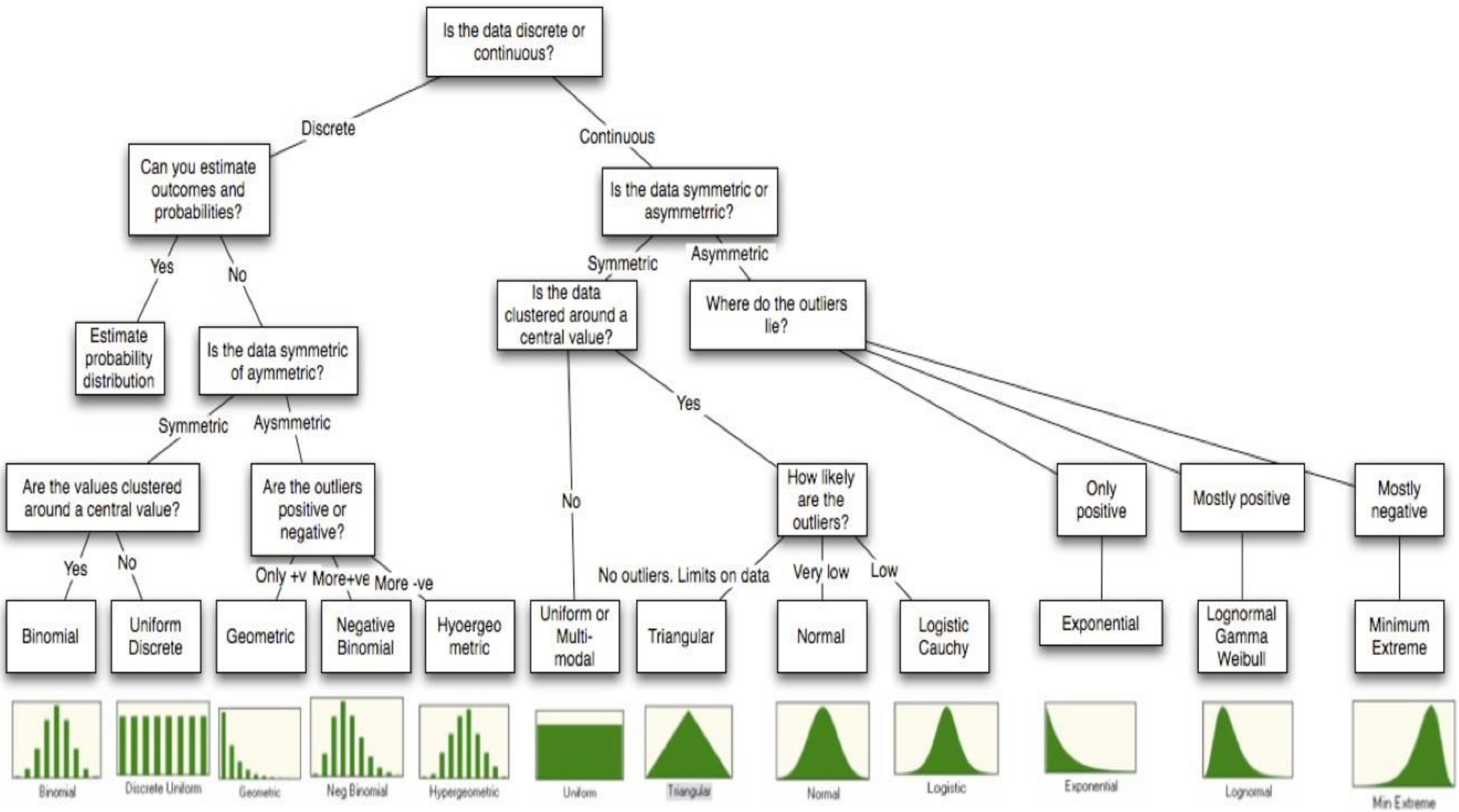
در این صورت می گوییم X دارای توزیع وایبول با پارامترهایλ و k می باشد.

از این توزیع اغلب برای مدلسازی زمان تا وقوع رخدادها در مواقعی که احتمال وقوع با زمان تغییر می کند استفاده می شود .(فرآیند های حافظه دار) در صورتیکه احتمال وقوع ثابت باشد بصورت توزیع نمایی تبدیل می گردد.





Distributional Choices



منابع:

- pages.stern.nyu.edu/~adamodar/New_Home_Page/StatFile/statdistns.htm
- http://www.epixanalytics.com/modelassist/AtRisk/Model_Assist.htm#Distributions/
- http://www.env.gov.bc.ca/epd/remediation/guidance/technical/pdf/12/gd04_all.pdf
- http://onlinestatbook.com/Online_Statistics_Education.htm
- http://www.cs.cmu.edu/~rbd/doc/nyquist/part14.html
- Discrete-Event System Simulation - Kellogg School of Management , Gigi Yuen