مروری بر برخی توزیع های آماری

دانشگاه مازندران

برخی توزیع های مهم گسسته: توزیع یکنواخت گسسته، توزیع برنولی، توزیع دوجمله ای، توزیع دوجمله ای منفی، توزیع هندسی و توزیع پواسن.

برخی توزیع های مهم پیوسته: توزیع یکنواخت پیوسته، توزیع نرمال، توزیع گاما، توزیع نمایی، توزیع کی–دو، توزیع بتا، توزیع وایبل و توزیع پارتو.

توزیع یکنواخت گسسته $(Discrete\ Uniform\ Distribution)$

اگر متغیر تصادفی X یکی از مقادیر $۱, ۲, ۳, \ldots, k$ را با احتمال های برابر قبول کند در آنصورت X را یکنواخت گسسته گویند که تابع احتمال آن بصورت زیر باشد

$$f_X(x) = P(X = x) = \frac{۱}{k} \qquad x = ۱, ۲, \ldots, k.$$

خواص.

$$if\ X \sim DU(\{۱, ۲, \ldots, k\}) \Rightarrow \begin{cases} E(X) = \frac{k+۱}{۲} \\ Var(X) = \frac{k^۲-۱}{۱۲} \\ M_X(t) = \frac{e^t(۱-e^{kt})}{k(۱-e^t)} \end{cases}$$

توزیع برنولی (Bernoulli Distribution)

متغیر تصادفی X توزیع برنولی دارد و می نویسیم $X \sim B(1, p)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x;p) = p^x\,(1-p)^{1-x} \qquad x = 0, 1, \qquad 0 < p \leq 1.$$

خواص.

$$if\ X \sim B(1,p) \Rightarrow \begin{cases} E(X) = p \\ Var(X) = p\,q \\ M_X(t) = p\,e^t + q \end{cases}$$

که در آن $q = 1 - p$.

توزیع دوجمله ای (Binomial Distribution)

متغیر تصادفی X توزیع دوجمله ای دارد و می نویسیم $X \sim B(n,\, p)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x;n,p) = \binom{n}{x} p^x\,(۱-p)^{n-x} \qquad x = ۰,\, ۱, \ldots, n, \qquad ۰ < p \leq ۱.$$

خواص.

$$if\ X \sim B(n,p) \Rightarrow \begin{cases} E(X) = np \\ Var(X) = np\,q \\ M_X(t) = (p\,e^t + q)^n \end{cases}$$

که در آن $q = ۱ - p$.

توزیع دوجمله ای منفی (Negative Binomial Distribution)

متغیر تصادفی X توزیع دوجمله ای منفی دارد و می نویسیم $X \sim NB(r,\ \theta)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x; r, \theta) = \binom{x+r-1}{x} \theta^x (1-\theta)^r,$$

که در آن $r > 0$, $0 < \theta \leq 1$, تعداد موفقیت ها $- \ x \in \{0,\ 1,\ 2,\ \ldots\}$.

خواص.

$$if\ X \sim NB(r,\theta) \Rightarrow \begin{cases} E(X) = \frac{r\,\theta}{(1-\theta)} \\ Var(X) = \frac{r\,\theta}{(1-\theta)^2} \\ M_X(t) = (\frac{1-\theta}{1-\theta\, e^t})^r \end{cases}$$

توزیع هندسی (مدل آزمایش) (Geometric Distribution)

متغیر تصادفی X توزیع هندسی دارد و می نویسیم $X \sim Ge(\theta)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x;\theta) = \theta\,(1-\theta)^{x-1}, \quad x = 1, 2, \ldots, \quad 0 < \theta \leq 1.$$

خواص.

$$if\ X \sim Ge(\theta) \Rightarrow \begin{cases} E(X) = \frac{1}{\theta} \\ Var(X) = \frac{1-\theta}{\theta^2} \\ M_X(t) = \frac{\theta\, e^t}{1-(1-\theta)\, e^t} \end{cases}$$

توزیع هندسی (مدل شکست)

متغیر تصادفی X توزیع هندسی (مدل شکست) دارد اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x;\theta) = \theta\,(۱-\theta)^x, \quad x = ۰, ۱, ۲, \ldots, \quad ۰ < \theta \leq ۱.$$

خواص.

$$if\ X \sim Ge^*(\theta) \Rightarrow \begin{cases} E(X) = \frac{۱-\theta}{\theta} \\ Var(X) = \frac{۱-\theta}{\theta^۲} \\ M_X(t) = \frac{\theta}{۱-(۱-\theta)\,e^t} \end{cases}$$

توزیع پواسن (Poisson Distribution)

متغیر تصادفی X توزیع پواسن با پارامتر λ دارد و می نویسیم $X \sim Pois(\lambda)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = ۰, ۱, ۲, \dots, \quad \lambda > ۰.$$

خواص.

$$if\ X \sim Pois(\lambda) \Rightarrow \begin{cases} E(X) = \lambda \\ Var(X) = \lambda \\ M_X(t) = e^{\lambda(e^t - ۱)} \end{cases}$$

تعریف توزیع یکنواخت پیوسته $(Uniform\ Distribution)$

گوییم متغیر تصادفی X دارای توزیع یکنواخت در فاصله $(a,\, b)$ است و می نویسیم $X \sim U(a,\, b)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x) = \frac{۱}{b-a}, \qquad a < x < b$$

خواص.

$$if\ X \sim U(a,\, b) \Rightarrow \begin{cases} E(X) = \frac{a+b}{۲} \\ Var(X) = \frac{(b-a)^{۲}}{۱۲} \\ M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)} \end{cases}$$

توزیع نرمال $(Normal\ Distribution)$

گوییم متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ است و می نویسیم $X \sim N(\mu, \sigma^۲)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x;\mu,\sigma^۲) = \frac{۱}{\sigma\sqrt{۲\pi}} e^{-\frac{۱}{۲}(\frac{x-\mu}{\sigma})^۲}, \qquad -\infty < x < \infty,$$

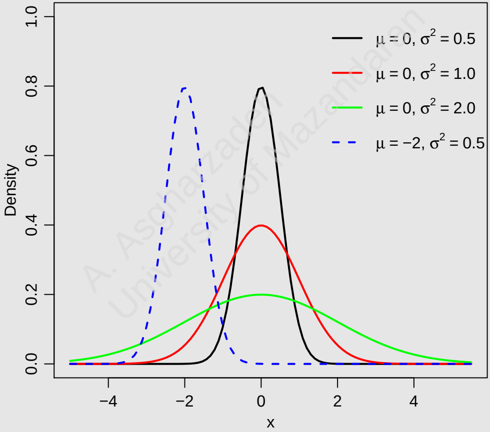
که در آن $\sigma > ۰,\ \mu \in (-\infty, \infty)$.

خواص.

$$if\ X \sim N(\mu,\sigma^۲) \Rightarrow \begin{cases} E(X) = \mu, \\ Var(X) = \sigma^۲ \\ M_X(t) = e^{\mu t + \frac{۱}{۲}\sigma^۲ t^۲} \end{cases}$$

توزیع نرمال استاندارد: توزیع نرمال با $\mu = ۰$ و $\sigma = ۱$، توزیع نرمال استاندارد نامیده می شود.

شکل: نمودار تابع چگالی احتمال توزیع نرمال به ازای مقادیر مختلف پارامترها



مثال ۱. متغیر تصادفی X توزیع نرمال با میانگین $\mu = ۱۵$ و واریانس $\sigma^۲ = ۱۲$ دارد.
مطلوب است

الف) $P(X < ۱۰)$
ب) $P(X > ۱۶)$
ج) $P(۱۲ < X < ۲۰)$

مثال ۱. متغیر تصادفی X توزیع نرمال با میانگین $\mu = ۱۵$ و واریانس $\sigma^۲ = ۱۲$ دارد. مطلوب است

الف) $P(X < ۱۰)$
ب) $P(X > ۱۶)$
ج) $P(۱۲ < X < ۲۰)$

مثال ۲. میزان رشد یک نوع نهال طی یک ماه، متغیر تصادفی نرمال با میانگین ۱۵ سانتی متر و انحراف معیار ۱/۲ سانتی متر است. احتمال اینکه نهالی از این نوع طی یک ماه بیش از ۱۷ سانتی متر رشد کند چقدر است؟ این احتمال در عمل به چه معناست؟ ۹۰٪ نهال ها حداکثر چه رشدی دارند؟

تعریف توزیع گاما $(Gamma\ Distribution)$

گوییم متغیر تصادفی X دارای توزیع گاما با پارامترهای α و β است و می نویسیم $X \sim \Gamma(\alpha, \beta)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} x^{\alpha-1}\, e^{-\frac{x}{\beta}}, \qquad x > 0,\ \alpha,\ \beta > 0$$

که در آن $\Gamma(\alpha)$ تابع گاما نام دارد که بصورت زیر تعریف می شود.

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1}\, e^{-x}\, \mathrm{d}x \qquad \alpha > 0$$

خواص

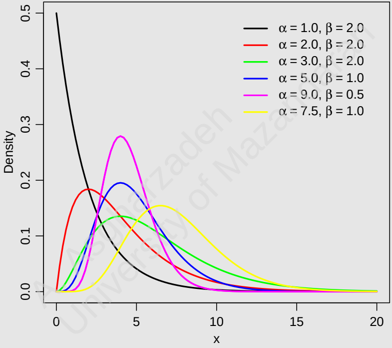
خواص تابع گاما.

$$\begin{aligned} ۱)\ \Gamma(۱) &= ۱ \\ ۲)\ \Gamma(\alpha+۱) &= \alpha\,\Gamma(\alpha) \\ ۳)\ \Gamma(n) &= (n-۱)! \qquad n \in \mathbb{Z}^{+} - \{۰\} \quad \mathbb{Z} \in \mathbb{N} \\ ۴)\ \Gamma(\tfrac{۱}{۲}) &= \sqrt{\pi} \end{aligned}$$

خواص X

$$if\ X \sim \Gamma(\alpha,\beta) \Rightarrow \begin{cases} E(X) = \alpha\,\beta \\ Var(X) = \alpha\,\beta^{۲} \\ M_X(t) = (۱-\beta\,t)^{-\alpha}, \quad t < \frac{۱}{\beta} \end{cases}$$

شکل: نمودار تابع چگالی احتمال توزیع گاما به ازای مقادیر مختلف پارامترها



با توجه به نمودار فوق، رفتار منحنی تابع چگالی توزیع گاما به ازای مقادیر مختلف پارامترها، نزولی و ابتدا صعودی سپس نزولی (صعودی-نزولی) می باشد.

مثال ۳. در شهری معین، مصرف روزانه برق را برحسب میلیون کیلووات ساعت می توان به صورت متغیری تصادفی در نظر گرفت که توزیع گاما با ۳ $= \alpha$ و ۲ $= \beta$ است. اگر کارخانه برق این شهر ظرفیتی برابر ۱۲ میلیون کیلووات ساعت داشته باشد، احتمال آنکه مقدار برق برای روز مفروضی ناکافی باشد چقدر است؟

توزیع نمایی $(Exponential\ Distribution)$

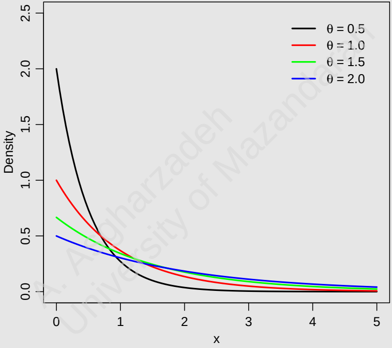
حالت خاص از توزیع گاما با $\alpha = ۱$ و $\beta = \theta$ را توزیع نمایی گویند. توزیع نمایی با پارامتر θ را با نماد $Exp(\theta)$ نمایش می دهند. لذا

$$f(x) = \frac{۱}{\theta} e^{-\frac{x}{\theta}}, \qquad x > ۰, \theta > ۰$$

خواص.

$$if\ X \sim Exp(\theta) = \Gamma(۱, \theta) \Rightarrow \begin{cases} E(X) = \theta \\ Var(X) = \theta^{۲} \\ M_X(t) = (۱ - \theta t)^{-۱}, \quad \theta < \frac{۱}{t} \end{cases}$$

شکل: نمودار تابع چگالی احتمال توزیع نمایی به ازای مقادیر مختلف θ



با توجه به نمودار، رفتار منحنی تابع چگالی نمایی به ازای مقادیر مختلف θ همواره نزولی می باشد.

یک ویژگی مهم متغیر تصادفی نمایی، فقدان حافظه (بی حافظه گی) آن است.

تعریف: ویژگی فقدان حافظه

گوییم متغیر تصادفی X (یا اصطلاحا توزیع X) فاقد حافظه است اگر

$$P(X > s + t | X > t) = P(X > s), \qquad t,\ s \geq ۰$$

تحت شرایط کلی، توزیع هندسی تنها توزیع گسسته فاقد حافظه، و توزیع نمایی تنها توزیع پیوسته فاقد حافظه است.

مثال ۴. طول عمر هر ذره از یک ایزوتوپ رادیواکتیو X (برحسب روز) متغیر تصادفی نمایی با میانگین ۴٫۲ روز است.

الف) تابع توزیع تجمعی X را به دست آورید.
ب) احتمال آنکه یک ذره از این ایزوتوپ بیش از ۷ روز عمر کند، چقدر است؟ نیم عمر این ایزوتوپ چقدر است؟

مثال ۵. در یک خط تولید، قطعاتی ساخته می شود که طول عمر آنها توزیع نمایی با میانگین ۷/۶ (هزار ساعت) است. احتمال آنکه قطعه ای از این نوع بیش از ۵ (هزار ساعت) کار کند، چقدر است؟

فرض کنید یک قطعه ۲ (هزار ساعت) کار کرده است، احتمال آنکه ۵ (هزار ساعت) دیگر کار کند چقدر است؟

توزیع کی دو یا $\chi^۲$ $(Chisqure)$

توزیع کی دو حالت خاصی از توزیع گاماست با پارامترهای $\alpha = \frac{n}{۲}$ و $\beta = ۲$. توزیع کی دو با n درجه آزادی را با نماد $\chi^۲(n)$ نمایش می دهند لذا

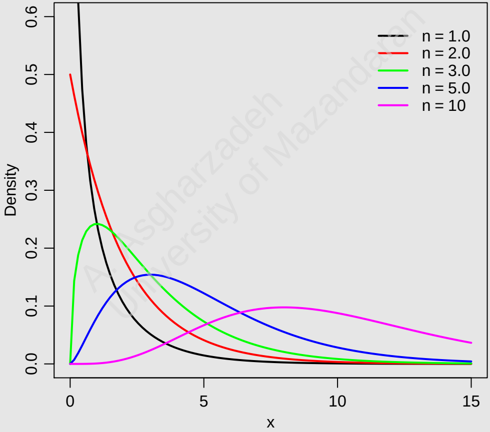
$$f(x) = \frac{۱}{۲^{\frac{n}{۲}}\Gamma(\frac{n}{۲})} x^{\frac{n}{۲}-۱} e^{-\frac{x}{۲}}, \qquad x > ۰$$

خواص.

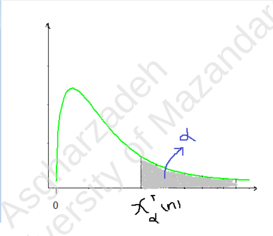
$$if\ X \sim \chi^۲(n) \Rightarrow \begin{cases} E(X) &= n \\ Var(X) &= ۲n \\ M_X(t) &= (۱-۲t)^{-\frac{n}{۲}} \end{cases}$$

مفهوم درجه آزادی: درجه آزادی یک مجموع تعداد متغیرهای مستقل آن مجموع است یعنی تعداد متغیرهایی که می توانند آزادانه مقادیر بگیرند (حرکت کنند).
(محدودیت) تعداد قیدها - تعداد متغیرهای آن مجموع = درجه آزادی یک مجموع

شکل: نمودار تابع چگالی احتمال توزیع کی دو به ازای مقادیر مختلف n



منحنی توزیع کی دو. نمودار تابع چگالی $\chi^{۲}(n)$ به صورت زیر می باشد که $\chi^{۲}_{\alpha}(n)$ نقطه ای روی محور طول هاست که سطح زیر منحنی چگالی $\chi^{۲}(n)$ در سمت راست این نقطه α شود.



$$P(\chi^{۲}(n) > \chi^{۲}_{\alpha}(n)) = \alpha$$

مقادیر $\chi^{۲}_{\alpha}(n)$ را می توان برای مقادیر مختلف n از روی جدول $\chi^{۲}$ پیدا کرد. به عنوان مثال $\chi^{۲}_{۰٫۰۱}(۱) = ۶٫۶۳۵$ یا $\chi^{۲}_{۰٫۹۵}(۱۰) = ۳٫۹۴$.

$$P(\chi^{۲}_{(۱۰)} > ۳٫۹۴) = P(\chi^{۲}_{(۱۰)} > \chi^{۲}_{۰٫۹۵}(۱۰)) = ۰٫۹۵$$

$$P(\chi^{۲}(۵) > ۱۲٫۸۳) = P(\chi^{۲}(۵) > \chi^{۲}_{۰٫۰۲۵}(۵)) = ۰٫۰۲۵$$

$$P(\chi^{۲}(۳) < ۱۱٫۳۴۵) = ۱ - P(\chi^{۲}(۳) > ۱۱٫۳۴۵)$$

$$= ۱ - P(\chi^{۲}(۳) > \chi^{۲}_{۰٫۰۱}(۳)) = ۱ - ۰٫۰۱ = ۰٫۹۹$$

توزیع بتا $(Beta\ Distribution)$

گوییم متغیر تصادفی X دارای توزیع بتا با پارامترهای α و β است و می نویسیم $X \sim Beta(\alpha,\, \beta)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x;\alpha,\beta) = \frac{x^{\alpha-۱}\,(۱-x)^{\beta-۱}}{B(\alpha,\,\beta)}, \qquad ۰ < x < ۱,\ \ \alpha,\,\beta > ۰$$

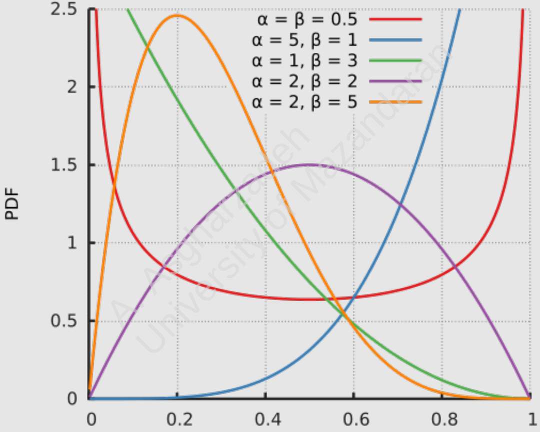
که در آن $B(\alpha,\,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

* اگر $\alpha = \beta = ۱$ یک توزیع یکنواخت در فاصله $[۰,۱]$ خواهیم داشت.

خواص.

$$if\ X \sim Beta(\alpha,\beta) \Rightarrow \begin{cases} E(X) = \frac{\alpha}{\alpha+\beta} \\ Var(X) = \frac{\alpha\,\beta}{(\alpha+\beta)^{۲}\,(\alpha+\beta+۱)} \end{cases}$$

شکل: نمودار تابع چگالی احتمال توزیع بتا به ازای مقادیر مختلف پارامترها



مثال ۶. اگر متغیر تصادفی X درصدی از درآمد افراد باشد که پس انداز می شود و فرض کنید توزیع آن بتا با پارامترهای $\alpha = ۲$ و $\beta = ۱۹$ باشد. احتمال اینکه فردی کمتر از ۵٪ درآمد خود را پس انداز کند چقدر است؟

مثال ۶. اگر متغیر تصادفی X درصدی از درآمد افراد باشد که پس انداز می شود و فرض کنید توزیع آن بتا با پارامترهای $\alpha = ۲$ و $\beta = ۱۹$ باشد. احتمال اینکه فردی کمتر از ۵٪ درآمد خود را پس انداز کند چقدر است؟

مثال ۷. درصدي از حشرات که بعد از نوعي سم پاشي کشته مي شوند متغير تصادفي بتا با ميانگين ۰/۶ و انحراف استاندارد ۰/۲ است احتمال اينکه دست کم نيمي از حشرات در اثر سم پاشي کشته شوند را محاسبه کنيد.

توزیع وایبل $(Weibull\ Distribution)$

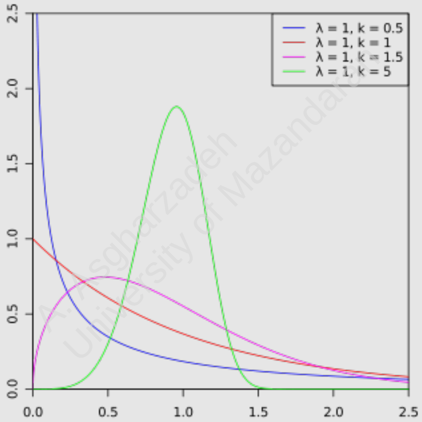
گوییم متغیر تصادفی X دارای توزیع وایبل با پارامترهای λ و k است و می نویسیم $X \sim We(\lambda, k)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x;\lambda,k) = \frac{k}{\lambda^k}\, x^{k-۱}\, e^{-(\frac{x}{\lambda})^k}, \qquad x > ۰,\ \lambda > ۰,\ k > ۰$$

خواص.

$$if\ X \sim We(\lambda,\ k) \Rightarrow \begin{cases} E(X) &= \lambda\,\Gamma(۱ + ۱/k) \\ Var(X) &= \lambda^۲\left[\Gamma(۱+\frac{۲}{k}) - \left(\Gamma(۱+\frac{۱}{k})\right)^۲\right] \end{cases}$$

شکل: نمودار تابع چگالی احتمال توزیع وایبل به ازای مقادیر مختلف پارامترها



مثال ۸. طول عمر یک قطعه اکترونیکی (هزار ساعت) توزیع وایبل با پارامترهای $k = ۲$ و $\lambda = ۵۰۰$ دارد.

الف) احتمال اینکه قطعه ای از این نوع بیش از ۳۰ (هزار ساعت) عمر کند چقدر است؟

ب) متوسط طول عمر این نوع قطعات چقدر است؟

توزیع پارتو $(Pareto\ Distribution)$

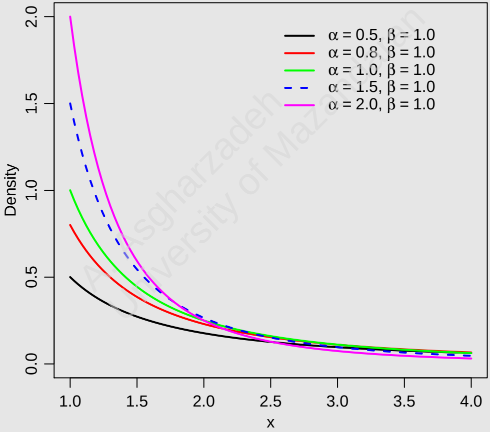
گوییم متغیر تصادفی X دارای توزیع پارتو با پارامترهای α و β است و می نویسیم $X \sim Pa(\alpha, \beta)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x; \alpha, \beta) = \frac{\alpha\, \beta^{\alpha}}{x^{\alpha+1}}, \qquad x > \beta,\ \alpha,\ \beta > 0$$

خواص.

$$if\ X \sim Pa(\alpha,\ \beta) \Rightarrow \begin{cases} E(X) = \frac{\alpha\,\beta}{\alpha-1}, & \alpha > 1 \\ Var(X) = \frac{\alpha\,\beta^{2}}{(\alpha-1)^{2}\,(\alpha-2)}, & \alpha > 2 \end{cases}$$

شکل: نمودار تابع چگالی احتمال توزیع پارتو به ازای مقادیر مختلف پارامترها



مثال ۹. فرض کنید متغیر تصادفی X دارای توزیع پارتو با پارامترهای $\alpha = ۲$ و $\beta = ۱$ است.

الف) هریک از مقدارهای $P(X < ۰/۵)$ و $P(۱ < X < ۲)$ را به دست آورید.

ب) برای هریک از مقدارهای $p = ۰/۸, ۰/۹۹$ مقدار x_p را طوری تعیین کنید که $P(X < x_p) = p$.