Estatísticas de ordem e outras distribuições importantes

Amostra Aleatória

Se $X_1, \dots, X_n$ são v.a.'s independentes e identicamente distribuídas (i.i.d.) com distribuição F, então $X_1, \dots, X_n$ são ditas uma *amostra aleatória (a.a.) de tamanho n* da *população* F.

Amostra Aleatória

Se $X_1, \ldots, X_n$ são v.a.'s independentes e identicamente distribuídas (i.i.d.) com distribuição F, então $X_1, \ldots, X_n$ são ditas uma *amostra aleatória (a.a.) de tamanho n* da *população* F.

Estatística

Seja $X_1, \ldots, X_n$ é uma a.a. de tamanho n de uma população e seja $T(x_1, \ldots, x_n)$ uma função tomando valores em $\mathbb{R}$. Então a variável aleatória $Y = T(X_1, \ldots, X_n)$ é dita uma *estatística* ou um *estimador pontual*. A distribuição de Y é chamada de *distribuição amostral de Y*.

Dada uma realização $X_1(\omega), \ldots, X_n(\omega)$ de uma a.a. de tamanho n, os dados ordenas de maneira crescente são denotados por $X_{(1,n)}, \ldots, X_{(n,n)}$, isto é, a sequência satisfaz

$$X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \ldots \leq X_{(n)}.$$

- $X_{(1)}$ é dita a k-ésima estatística de ordem de $X_1 \ldots, X_n$

Dada uma realização $X_1(\omega), \ldots, X_n(\omega)$ de uma a.a. de tamanho n, os dados ordenas de maneira crescente são denotados por $X_{(1,n)}, \ldots, X_{(n,n)}$, isto é, a sequência satisfaz

$$X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \ldots \leq X_{(n)}.$$

- $X_{(1)}$ é dita a k-ésima estatística de ordem de $X_1 \ldots, X_n$
- Note que em particular:

$$X_{(1)} = min\{X_1, \ldots, X_n\}$$
$$X_{(n)} = max\{X_1, \ldots, X_n\}$$

Teorema

Dadas $X_1, \ldots, X_n$ v.a.'s independentes com funções de distribuição $F_1, \ldots, F_n$ respectivamente, a distribuições de $X_{(1)}$ e $X_{(n)}$ são dadas respectivamente por:

$$F_{(1)}(z) = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(z)] \quad e \quad F_{(n)}(z) = \prod_{i=1}^{n} F_{X_i}(z)$$

Teorema

Dadas $X_1, \ldots, X_n$ v.a.'s independentes com funções de distribuição $F_1, \ldots, F_n$ respectivamente, a distribuições de $X_{(1)}$ e $X_{(n)}$ são dadas respectivamente por:

$$F_{(1)}(z) = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(z)] \quad e \quad F_{(n)}(z) = \prod_{i=1}^{n} F_{X_i}(z)$$

Colorário

Se $X_1, \ldots, X_n$ são i.i.d. com distribuição F

$$F_{(1)}(z) = 1 - [1 - F(z)]^n \qquad e \qquad F_{(n)}(z) = [F(z)]^n$$

Nesse caso, suas respectivas f.d.p são dadas por

$$f_{(1)}(z) = n[1 - F(z)]^{n-1} f(z) \qquad e \qquad f_{(n)}(z) = n[F(z)]^{n-1} f(z)$$

1. Seja $X \sim Exp(2)$, $Y = max\{X, 2\}$ e $W = min\{X, 2\}$. Determine a função de distribuição de Y e W.

2. Seja X uma v.a. com f.d.p

$$f(x) = \begin{cases} \frac{x}{4} & se\ 0 \leq x < 2 \\ \frac{1}{8} & se\ 2 \leq x < 6 \\ 0 & c.\ c. \end{cases}$$

   Determine a função de distribuição de $Y = \min\{X, 3\}$

De forma geral

Teorema

Dadas $X_1, \dots, X_n$ v.a.'s independentes com f.d.p. f a f.d.p de $X_{(k)}$ e a f.d.p conjunta de $X_{(1)}, \dots, X_{(n)}$ são dadas respectivamente por:

$$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!}[1 - F(x)]^{n-k} F(x)^{k-1} f(x)$$

$$f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots, x_n) = n! f(x_1) \dots f(x_n)$$

Além disso para qualquer par de v.a's $X_{(i)}, \dots, X_{(j)}$ a sua conjunta $f_{(k)}(x_i, x_j)$ é dada por

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(x_i)^{i-1} [F(x_j) - F(x_i)]^{j-i-1} [1 - F(x_j)]^{n-j} f(x_i) f(x_j)$$

para $x_i < x_j$

3. Sejam $X_1, X_2, \ldots, X_n$ v.a. i.i.d com distribuição uniforme no intervalo $(0, 1)$. Determine a f.d.p conjunta de $X_{(1)}$ e $X_{(n)}$

Mediana

Dada $X_1, \ldots, X_n$ de uma a.a. de tamanho n, a *mediana* da amostra é definida como

$$Md = \begin{cases} X_{(\frac{n+1}{2})} & se \ n = 2k+1 \\ \dfrac{X_{(\frac{n}{2})} + X_{(\frac{n+1}{2})}}{2} & se \ n = 2k \end{cases}$$

4. Sejam $X_1, X_2, \ldots, X_7$ v.a. i.i.d com distribuição uniforme no intervalo $(0, 1)$. Determine a f.d.p da mediana da amostra

Outras distribuições importantes

Distribuição Gama: $X \sim Gama(k, \theta)$

- **Função densidade de probabilidade**: para $k > 0$ e $\theta > 0$

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \; x > 0$$

ou com $r = k$ e $\beta = 1/\theta$

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

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

  - $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$
  - Se k é inteiro $\Gamma(k) = (k-1)\Gamma(k-1) \implies \Gamma(k) = (k-1)!$
  - $\Gamma(1/2) = \sqrt{\pi}$

Distribuição Gama: $X \sim Gama(k, \theta)$

- **Função densidade de probabilidade**: para $k > 0$ e $\theta > 0$

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \ x > 0$$

  ou com $r = k$ e $\beta = 1/\theta$

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

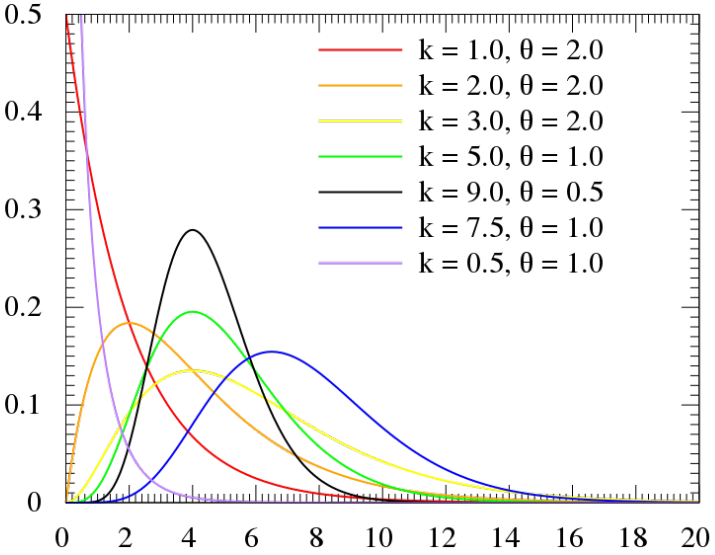
  onde $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx, k > 0$

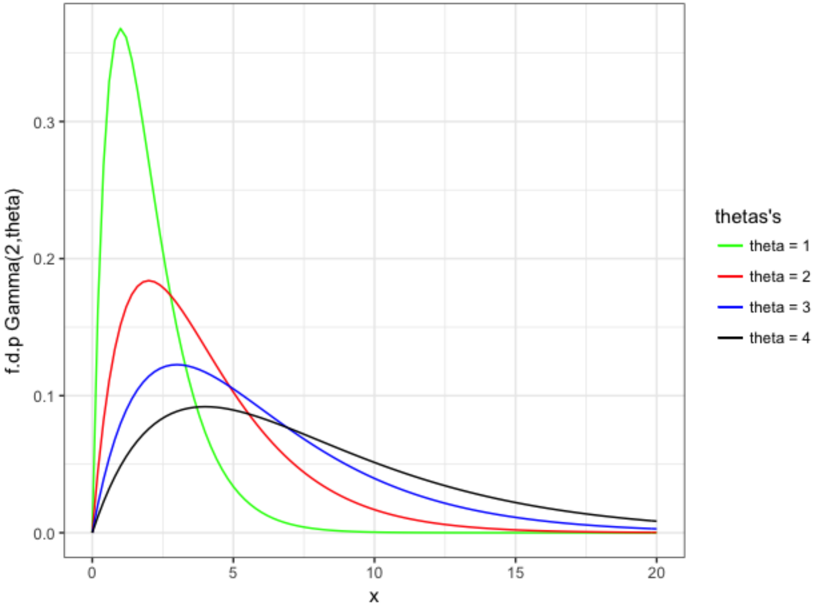
  - $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$
  - Se k é inteiro $\Gamma(k) = (k-1)\Gamma(k-1) \implies \Gamma(k) = (k-1)!$
  - $\Gamma(1/2) = \sqrt{\pi}$

- A distribuição Gama é uma generalização da exponencial

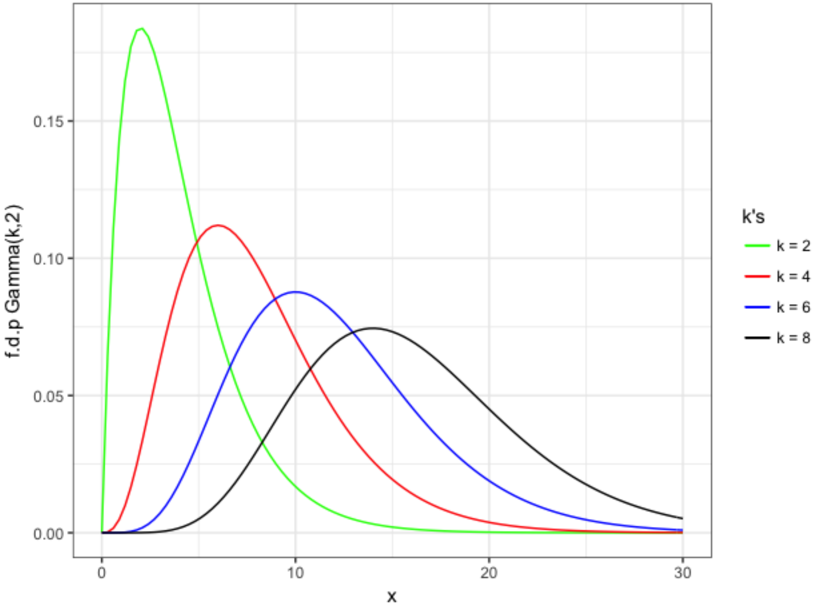
$$X \sim Gamma(1, \theta) \implies X \sim Exp(1/\theta)$$

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \; x > 0$$





f.d.p Gamma(2,theta)
0.0
0.1
0.2
0.3
0
5
10
15
20
x
thetas's
theta = 1
theta = 2
theta = 3
theta = 4


f.d.p Gamma(k,2)
0.15
0.10
0.05
0.00
0
10
20
30
x
k's
k = 2
k = 4
k = 6
k = 8

Distribuição Gama: $X \sim Gama(k, \theta)$

- **Função densidade de probabilidade**: para $\theta, k > 0$

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad x > 0, \; com \; \Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$$

- **Valor Esperado**: $EX = k\theta$
- **Variância**: $Var(X) = k\theta^2$

Distribuição Gama: $X \sim Gama(k, \theta)$

- **Função densidade de probabilidade**: para $\theta, k > 0$

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad x > 0, \; com \; \Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$$

- **Valor Esperado**: $EX = k\theta$
- **Variância**: $Var(X) = k\theta^2$
- **Função geradora de momentos**: $M(t) = (1 - \theta t)^{-k}, \quad t < 1/\theta$

Propriedades

- Se $X_1, \ldots, X_n$ são v.a. i.i.d. com distribuição $Gama(k_i, \theta)$, $i = 1, \ldots, n$, então

$$\sum_{i=1}^{n} X_i \sim Gama(\sum_{i=1}^{n} k_i, \theta)$$

- $X \sim Gama(k, \theta) \Longrightarrow cX \sim Gama(k, c\theta)$

ma

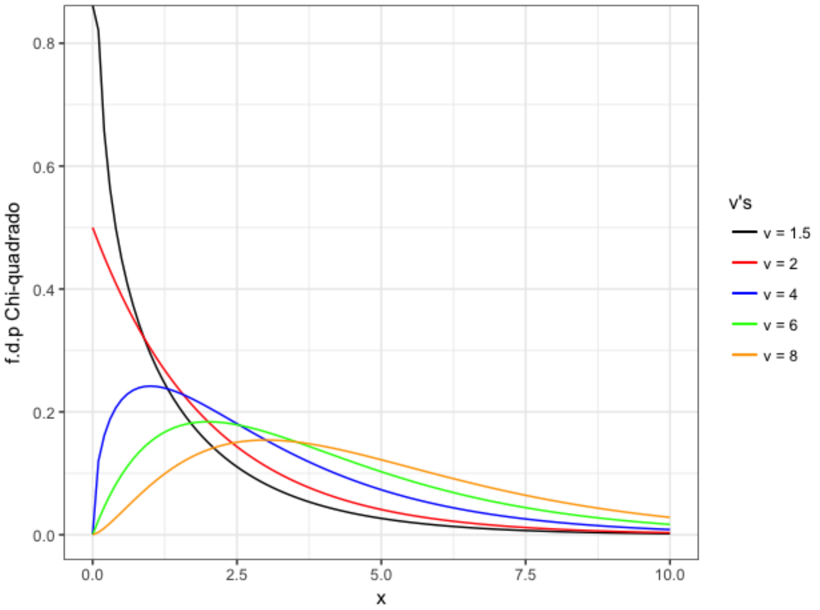
- Se $X_1, \ldots, X_n$ são v.a. i.i.d com distribuição $Exp(\lambda)$ então

$$\sum_{i=1}^{n} X_i \sim Gama(n, 1/\lambda)$$

Chi-quadrado: $X \sim \chi^2_\nu$

- Caso particular de uma Gama$(\nu/2, 2)$, $\nu = 1, 2, \ldots$
- **Função densidade de probabilidade**:

$$f(x) = \frac{1}{\Gamma(\nu/2) 2^{\nu/2}} \, x^{\nu/2-1} e^{-x/2}, \; x > 0$$


0.8
0.6
0.4
0.2
0.0
f.d.p Chi-quadrado
0.0
2.5
5.0
7.5
10.0
x
v's
v = 1.5
v = 2
v = 4
v = 6
v = 8

Chi-quadrado: $X \sim \chi^2_\nu$

- Caso particular de uma Gama$(\nu/2, 2)$, $\nu = 1, 2, \ldots$
- **Função densidade de probabilidade**:

$$f(x) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}}\, x^{\nu/2-1} e^{-x/2}, \; x > 0$$

- **Valor Esperado**: $EX = \nu$
- **Variância**: $Var(X) = 2\nu$
- **Função geradora de momentos**: $M(t) = (1-2t)^{-\nu/2}$, $t < 1/2$

Propriedades

- Se $Z_1, \ldots, Z_n$ são v.a. i.i.d com distribuição $N(0, 1)$ então

$$\sum_{i=1}^{n} Z_i^2 \sim \chi_\nu^2$$

Propriedades

- Se $Z_1, \ldots, Z_n$ são v.a. i.i.d com distribuição $N(0,1)$ então

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_\nu$$

- Se $X_1, \ldots, X_n$ são v.a.'s independes com $X_i \sim \chi^2_{\nu_i}$, então

$$\sum_{i=1}^{n} X_i \sim Gama\left(\frac{1}{2}\sum_{i=1}^{n} \nu_i, 2\right)$$

t-student: $T \sim t_\nu$

- Sejam Z e Y v.a.'s independentes com $Z \sim N(0,1)$ e $Y \sim \chi^2_\nu$

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

O parâmetro ν é dito *graus de liberdade* da distribuição t.

t-student: $T \sim t_\nu$

- Sejam Z e Y v.a.'s independentes com $Z \sim N(0,1)$ e $Y \sim \chi^2_\nu$

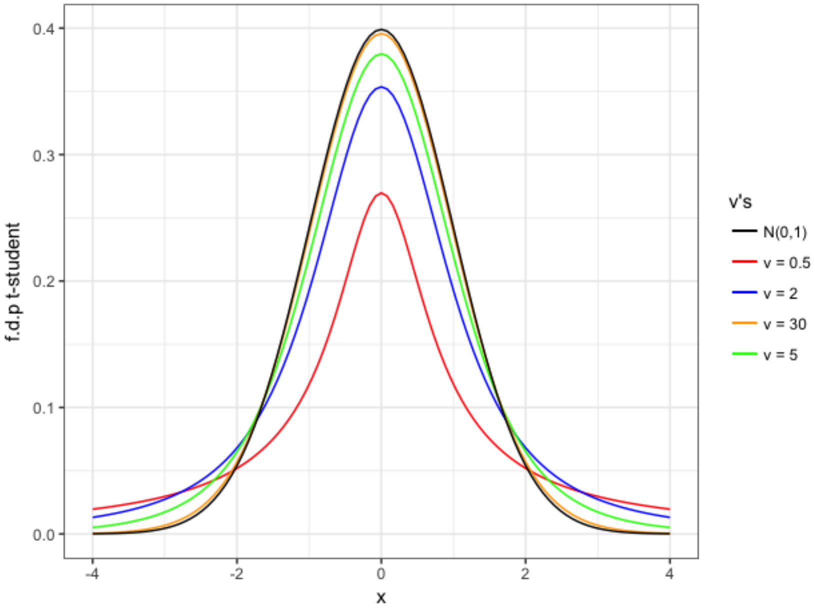
$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

O parâmetro ν é dito *graus de liberdade* da distribuição t.

- **Função densidade de probabilidade**:

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}(1+t^2/\nu)^{-(\nu+1)/2}, \ -\infty < t < \infty.$$

fdp



Propriedades

- Sejam Z e Y v.a.'s independentes com $Z \sim N(0,1)$ e $Y \sim \chi^2_\nu$

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

- **Função densidade de probabilidade**:

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}(1 + t^2/\nu)^{-(\nu+1)/2}, \; -\infty < t < \infty.$$

- **Valor Esperado**: $EX = 0, \; \nu > 1$
- **Variância**: $Var(X) = \frac{\nu}{\nu-2}, \nu > 2$ e ∞ para $1 < \nu \leq 2$

Propriedades

- Sejam Z e Y v.a.'s independentes com $Z \sim N(0,1)$ e $Y \sim \chi^2_\nu$

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

- **Função densidade de probabilidade**:

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}(1+t^2/\nu)^{-(\nu+1)/2}, \ -\infty < t < \infty.$$

- **Valor Esperado**: $EX = 0, \ \nu > 1$
- **Variância**: $Var(X) = \frac{\nu}{\nu-2}, \nu > 2$ e ∞ para $1 < \nu \leq 2$
- Se $X \sim t_\nu$ com $\nu > 30$, então X pode ser aproximada por $N(0,1)$

F-Snedecor: $W \sim F(\nu_1, \nu_2)$

- Dadas X e Y v.a.'s independentes com $X \sim \chi^2_{\nu_1}$ e $Y \sim \chi^2_{\nu_2}$ então

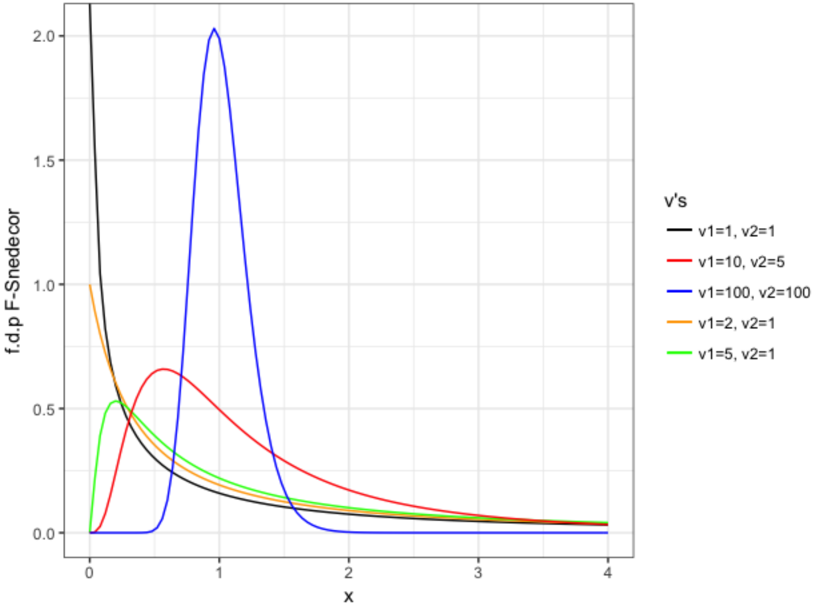
$$W = \frac{X/\nu_1}{Y/\nu_2} \sim F(\nu_1, \nu_2)$$

- **Função densidade de probabilidade**:

$$f(w) = \frac{\Gamma((\nu_1+\nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{w^{(\nu_1-2)/2}}{(1+\nu_1 w/\nu_2)^{(\nu_1+\nu_2)/2}}, \quad w > 0$$

- **Valor Esperado**: $EW = \dfrac{\nu_2}{\nu_1 - 2}$

- **Variância**: $Var(W) = \dfrac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$



f.d.p F-Snedecor
x
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
v's
v1=1, v2=1
v1=10, v2=5
v1=100, v2=100
v1=2, v2=1
v1=5, v2=1

Propriedades

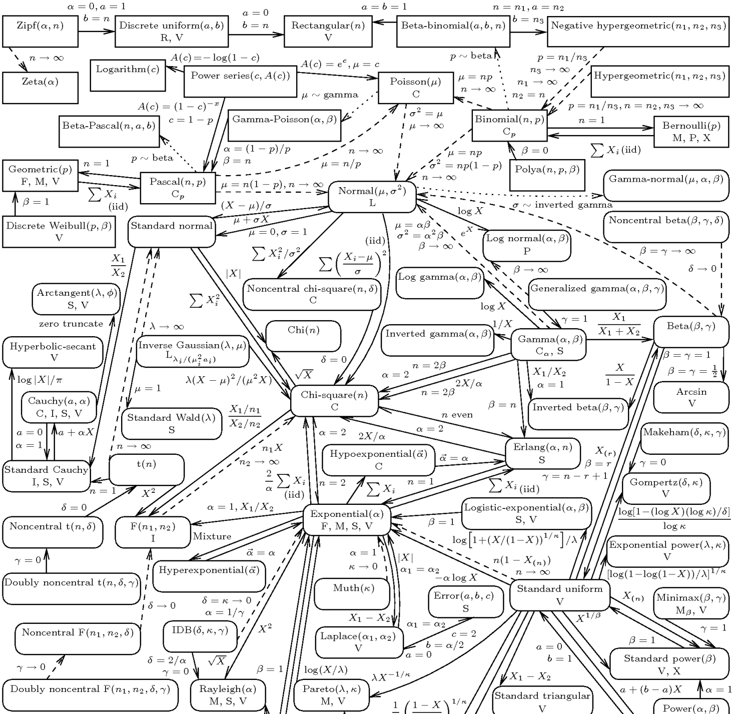
- Se $W_i \sim Gama(k_i, \theta_i), i = 1, 2$ independentes então

$$\frac{k_2\theta_2 W_1}{k_1\theta_1 W_2} \sim F(2k_1, 2k_2)$$

- Se $X \sim F(d_1/2, d_2/2)$ então

$$\frac{d_1 X/d_2}{1 + d_1 X/d_2} \sim Beta(d_1/2, d_2/2)$$

- Se $W \sim F(\nu_1, \nu_2)$ e $Y = lim_{\nu_2 \to \infty} \nu_1 W$ então $Y \sim \chi^2_{\nu_1}$
- Se $W \sim F(\nu_1, \nu_2)$ então $X^{-1} \sim F(\nu_2, \nu_1)$
- Se $T \sim t_\nu$ então $T^2 \sim F(1, \nu)$ e $T^{-2} \sim F(\nu, 1)$



Zipf(α, n)
$\alpha = 0, a = 1$
$b = n$
Discrete uniform(a, b) R, V
$a = 0$
$b = n$
Rectangular(n) V
$a = b = 1$
Beta-binomial(a, b, n)
$n = n_1, a = n_2$
$b = n_3$
Negative hypergeometric(n_1, n_2, n_3)
$n \to \infty$
Zeta(α)
Logarithm(c)
$A(c) = -\log(1-c)$
Power series$(c, A(c))$
$A(c) = e^c, \mu = c$
Poisson(μ) C
$p \sim$ beta
$p = n_1/n_3$
$n_3 \to \infty$
$n_1 \to \infty$
$n_2 = n$
Hypergeometric(n_1, n_2, n_3)
$\mu \sim$ gamma
$\mu = np$
$n \to \infty$
$A(c) = (1-c)^{-x}$
$c = 1 - p$
Gamma-Poisson(α, β)
$\sigma^2 = \mu$
$\mu \to \infty$
Binomial(n, p) C_p
$p = n_1/n_3, n = n_2, n_3 \to \infty$
$n = 1$
Bernoulli(p) M, P, X
$\sum X_i$ (iid)
Beta-Pascal(n, a, b)
$p \sim$ beta
$\alpha = (1-p)/p$
$\beta = n$
$n \to \infty$
$\mu = n/p$
$\mu = np$
$\sigma^2 = np(1-p)$
$n \to \infty$
$\beta = 0$
Polya(n, p, β)
Geometric(p) F, M, V
$n = 1$
$\sum X_i$ (iid)
Pascal(n, p) C_p
$\mu = n(1-p), n \to \infty$
Normal(μ, σ^2) L
Gamma-normal(μ, α, β)
$\sigma \sim$ inverted gamma
$\beta = 1$
Discrete Weibull(p, β) V
$(X-\mu)/\sigma$
$\mu + \sigma X$
$\mu = 0, \sigma = 1$
Standard normal
$\log X$
e^X
Noncentral beta(β, γ, δ)
$\mu = \alpha\beta$
$\sigma^2 = \alpha^2\beta$
$\beta \to \infty$
Log normal(α, β) P
$\beta = \gamma \to \infty$
$\delta \to 0$
(iid)
$\sum X_i^2/\sigma^2$
$\sum \left(\frac{X_i - \mu}{\sigma}\right)^2$
$\frac{X_1}{X_2}$
$|X|$
Arctangent(λ, ϕ) S, V
$\sum X_i^2$
Noncentral chi-square(n, δ) C
Log gamma(α, β)
$\beta \to \infty$
Generalized gamma(α, β, γ)
$\log X$
zero truncate
$\lambda \to \infty$
Chi(n)
$1/X$
$\gamma = 1$
$\frac{X_1}{X_1 + X_2}$
Beta(β, γ)
Hyperbolic-secant V
Inverse Gaussian(λ, μ) $L_{\lambda_i/(\mu_i^2 a_i)}$
Inverted gamma(α, β)
Gamma(α, β) C_α, S
$\beta = \gamma = 1$
$\delta = 0$
$\beta = \gamma = \frac{1}{2}$
$\log |X|/\pi$
$\lambda(X-\mu)^2/(\mu^2 X)$
$\sqrt{X}$
$\alpha = 2$
$n = 2\beta$
X_1/X_2
$\frac{X}{1-X}$
$\alpha = 1$
Arcsin V
$\mu = 1$
$2X/\alpha$
$n = 2\beta$
Cauchy(a, α) C, I, S, V
Standard Wald(λ) S
X_1/n_1
X_2/n_2
Chi-square(n) C
$\beta = n$
Inverted beta(β, γ)
n even
$a = 0$
$\alpha = 1$
$a + \alpha X$
$n \to \infty$
$\alpha = 2$
$\alpha = 2$
$2X/\alpha$
Makeham(δ, κ, γ)
$n_1 X$
$n_2 \to \infty$
Hypoexponential$(\vec{\alpha})$ C
$\vec{\alpha} = \alpha$
Erlang(α, n) S
$X_{(r)}$
$\beta = r$
$\gamma = 0$
Standard Cauchy I, S, V
t(n)
$\frac{2}{\alpha} \sum X_i$
(iid)
$n = 2$
$\sum X_i$
$n = 1$
$\gamma = n - r + 1$
$\sum X_i$ (iid)
Gompertz(δ, κ) V
$n = 1$
X^2
$\delta = 0$
$\alpha = 1, X_1/X_2$
Exponential(α) F, M, S, V
Logistic-exponential(α, β) S, V
$\log[1-(\log X)(\log \kappa)/\delta]$
$\log \kappa$
Noncentral t(n, δ)
F(n_1, n_2) I
Mixture
$\beta = 1$
$\log\left[1+(X/(1-X))^{1/\kappa}\right]/\lambda$
Exponential power(λ, κ) V
$\gamma = 0$
$\vec{\alpha} = \alpha$
Hyperexponential$(\vec{\alpha})$
$\alpha = 1$
$\kappa \to 0$
$|X|$
$\alpha_1 = \alpha_2$
$n(1 - X_{(n)})$
$n \to \infty$
$[\log(1-\log(1-X))/\lambda]^{1/\kappa}$
Doubly noncentral t(n, δ, γ)
Muth(κ)
$-\alpha \log X$
Error(a, b, c) S
Standard uniform V
$X_{(n)}$
Minimax(β, γ) M_β, V
$\delta \to 0$
$\delta = \kappa \to 0$
$\alpha = 1/\gamma$
$X_1 - X_2$
$X^{1/\beta}$
$\gamma = 1$
Noncentral F(n_1, n_2, δ)
IDB(δ, κ, γ)
X^2
Laplace(α_1, α_2) V
$\alpha_1 = \alpha_2$
$c = 2$
$b = \alpha/2$
$a = 0$
$a = 0$
$b = 1$
$\beta = 1$
Standard power(β) V, X
$\gamma \to 0$
$\delta = 2/\alpha$
$\gamma = 0$
$\sqrt{X}$
$\beta = 1$
$\log(X/\lambda)$
$\lambda X^{-1/\kappa}$
$X_1 - X_2$
$a + (b-a)X$
$\alpha = 1$
Doubly noncentral F$(n_1, n_2, \delta, \gamma)$
Rayleigh(α) M, S, V
Pareto(λ, κ) M, V
$\frac{1}{\lambda}\left(\frac{1-X}{X}\right)^{1/\kappa}$
Standard triangular V
Power(α, β)