

RandLib documentation

Aleksandr Samarin

October 22, 2018

Contents

I	General information	4
1	Calculation of sample moments	4
II	Continuous univariate distributions	5
2	Beta distribution	5
2.1	Arcsine distribution	7
2.2	Balding-Nichols distribution	8
2.3	Uniform distribution	8
3	Beta-prime distribution	10
4	Exponentially-modified Gaussian distribution	12
5	F-distribution	13
6	Gamma distribution	14
6.1	Chi-squared distribution	16
6.2	Erlang distribution	16
6.3	Exponential distribution	16
7	Geometric Stable distribution	17
7.1	Asymmetric Laplace distribution	17
7.2	Laplace distribution	17
8	Kolmogorov-Smirnov distribution	18
9	Logistic distribution	19

10 Log-normal distribution	20
11 Marchenko-Pastur distribution	21
12 Nakagami distribution	22
12.1 Chi distribution	22
12.2 Maxwell-Boltzmann distribution	23
12.3 Rayleigh distribution	23
13 Noncentral Chi-Squared distribution	24
14 Pareto distribution	25
15 Planck distribution	28
16 Stable distribution	29
16.1 Cauchy distribution	29
16.2 Levy distribution	29
16.3 Normal distribution	30
16.4 Holtsmark distribution	31
16.5 Landau distribution	31
17 Weibull	32
III Discrete univariate distributions	34
18 Beta-binomial distribution	34
19 Binomial distribution	35
19.1 Bernoulli	36
20 Hypergeometric distribution	37
21 Negative-Binomial (Polya) distribution	38
21.1 Geometric distribution	38
21.2 Pascal distribution	38
22 Poisson distribution	39
23 Skellam distribution	41
24 Uniform discrete distribution	42
25 Yule distribution	43
26 Zeta distribution	44

27 Zipf distribution	45
IV Bivariate distributions	46
28 Bivariate Normal distribution	46
29 Normal-Inverse-Gamma distribution	46
30 Trinomial distribution	46
V Circular distributions	47
31 von Mises distribution	47
32 Wrapped Exponential distribution	47
VI Singular distributions	48
33 Cantor distribution	48

Part I

General information

1 Calculation of sample moments

We use extension of Welford's method from Knuth. For every n -th element x we have

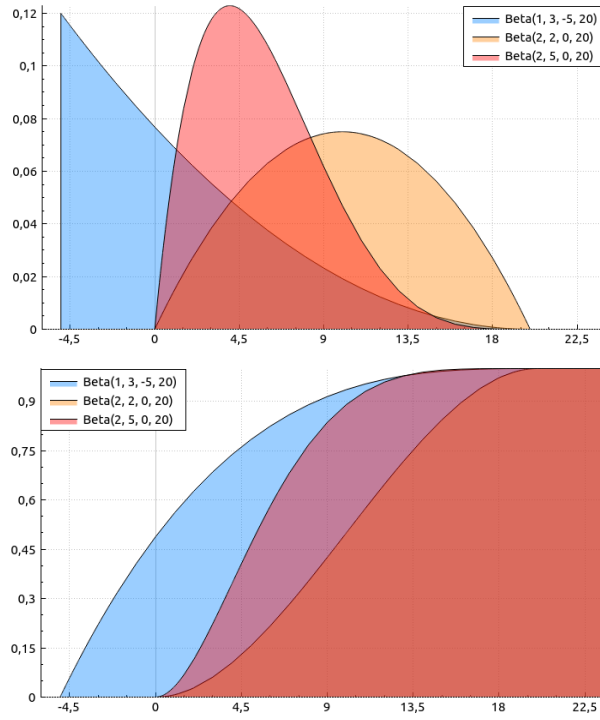
$$\begin{aligned}\delta &= x - m_1, \\ m'_1 &= m_1 + \frac{\delta}{n}, \\ m'_2 &= m_2 + \delta^2 \frac{n-1}{n}, \\ m'_3 &= m_3 + \delta^3 \frac{(n-1)(n-2)}{n^2} - 3\delta \frac{m_2}{n}, \\ m'_4 &= m_4 + \delta^4 \frac{(n-1)(n^2-3n+3)}{n^3} + 6\delta^2 \frac{m_2}{n^2} - 4\delta \frac{m_3}{n}.\end{aligned}$$

Then m'_1 , $\frac{m_2}{n}$, $\text{Skew}(X) = \frac{\sqrt{n}m'_3}{m_2^{3/2}}$ and $\text{Kurt}(X) = \frac{nm'_4}{m_2^2}$ (we return excess kurtosis).

Part II

Continuous univariate distributions

2 Beta distribution



Notation	$X \sim \mathcal{B}(\alpha, \beta, a, b)$ $X \sim \mathcal{B}(\alpha, \beta)$ with $a = 0, b = 1$
Parameters	$\alpha, \beta > 0, a, b \in \mathbb{R}$
Domain	$x \in [a, b]$
$f(x)$	$\frac{y^{\alpha-1}(1-y)^{\beta-1}}{(b-a)B(\alpha, \beta)}$ with $y = \frac{x-a}{b-a}$
$F(x)$	$I_y(\alpha, \beta)$ for $y = \frac{x-a}{b-a}$
$\mathbb{E}[X]$	$a + (b-a)\frac{\alpha}{\alpha+\beta}$
$\text{Var}(X)$	$(b-a)^2 \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Median	Searched numerically
Mode	$a + (b-a)\frac{\alpha-1}{\alpha+\beta-2}$ for $\alpha, \beta > 1$.
$\phi(t)$	Calculated numerically

Search of the median. In general, the value of median is unknown and searched numerically with initial value:

$$m \approx a + (b-a) \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}}$$

for $\alpha, \beta \geq 1$. However, there are analytical solutions for some particular values:

- $m = \frac{a+b}{2}$, for $\alpha = \beta$,
- $m = a + (b-a)(1 - 2^{-\frac{1}{\beta}})$, for $\alpha = 1$,
- $m = a + (b-a)2^{-\frac{1}{\alpha}}$, for $\beta = 1$.

Calculation of characteristic function. For $\alpha, \beta \geq 1$ we use numerical integration by definition

$$\phi(t) = \int_a^b \cos(tx) f(x) dx + i \int_a^b \sin(tx) f(x) dx.$$

For shape parameters < 1 , $f(x)$ has singularity points at 0 or 1 or both of them, and numerical integration is impossible. Then we use the following technique: firstly, we can show that

$$\phi(t|a, b) = \mathbb{E}[e^{it(a+(b-a)X)}] = e^{ita} \phi(z|0, 1)$$

with $z = (b - a)t$. Hence, w.l.o.g. we can consider standard case $a = 0, b = 1$. Then

$$\begin{aligned} \Re(\phi(z)) &= \frac{1}{B(\alpha, \beta)} \int_0^1 \cos(zx) x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 (\cos(zx) - 1) x^{\alpha-1} (1-x)^{\beta-1} dx + 1 \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 \frac{(\cos(zx) - 1) x^{\alpha-1} - (\cos(z) - 1)}{(1-x)^{1-\beta}} dx + 1 + \frac{\cos(z) - 1}{bB(\alpha, \beta)}. \end{aligned}$$

The integrand now doesn't have any singularities, neither for $\alpha < 1$, nor for $\beta < 1$. Analogously we transform the imaginary part:

$$\begin{aligned} \Im(\phi(z)) &= \frac{1}{B(\alpha, \beta)} \int_0^1 \sin(zx) x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 \frac{\sin(zx) x^{\alpha-1} - \sin(z)}{(1-x)^{1-\beta}} dx + \frac{\sin(z)}{bB(\alpha, \beta)}. \end{aligned}$$

Estimation of shapes with known support. Assume that $a = 0, b = 1$ and we have a sample $X = (X_1, \dots, X_n)$. Then a log-likelihood function is

$$\begin{aligned} \ln \mathcal{L}(\alpha, \beta | X) &= \sum_{i=1}^n \ln f(X_i; \alpha, \beta) \\ &= (\alpha - 1) \sum_{i=1}^n \ln X_i + (\beta - 1) \sum_{i=1}^n \ln(1 - X_i) - n \ln B(\alpha, \beta). \end{aligned} \tag{1}$$

Differentiating with respect to the shapes, we obtain

$$\begin{aligned} \frac{\partial \ln \mathcal{L}(\alpha, \beta | X)}{\partial \alpha} &= \sum_{i=1}^n \ln X_i + n(\psi(\alpha + \beta) - \psi(\alpha)), \\ \frac{\partial \ln \mathcal{L}(\alpha, \beta | X)}{\partial \beta} &= \sum_{i=1}^n \ln(1 - X_i) + n(\psi(\alpha + \beta) - \psi(\beta)). \end{aligned}$$

Differentiating again we get the Hessian matrix:

$$H(\ln \mathcal{L}(\alpha, \beta|X)) = n \cdot \begin{pmatrix} \psi_1(\alpha + \beta) - \psi_1(\alpha) & \psi_1(\alpha + \beta) \\ \psi_1(\alpha + \beta) & \psi_1(\alpha + \beta) - \psi_1(\beta) \end{pmatrix}.$$

Then we can find the estimators numerically, using Newton's procedure. The initial values of estimators are found via method of moments:

$$\hat{\alpha}_0 = \bar{X}_n \left(\frac{\bar{X}_n(1 - \bar{X}_n)}{\hat{s}_n^2} - 1 \right),$$

$$\hat{\beta}_0 = (1 - \bar{X}_n) \left(\frac{\bar{X}_n(1 - \bar{X}_n)}{\hat{s}_n^2} - 1 \right).$$

These values are applicable only if $\hat{s}_n^2 < \bar{X}_n(1 - \bar{X}_n)$. If this condition is not satisfied, we set $\hat{\alpha}_0 = \hat{\beta}_0 = 0.001$.

In the general case, when $a \neq 0$ or $b \neq 1$, we use the following transformation:

$$Y_i = \frac{X_i - a}{b - a}$$

and estimate parameters, using sample Y .

2.1 Arcsine distribution

Notation:

$$X \sim \text{Arcsine}(\alpha).$$

Relation to Beta distribution:

$$X \sim \mathcal{B}(1 - \alpha, \alpha, a, b).$$

Estimation of shape. For Arcsine distribution log-likelihood function (1) turns into

$$\ln \mathcal{L}(\alpha|X) = -\alpha \sum_{i=1}^n \ln X_i + (\alpha - 1) \sum_{i=1}^n \ln(1 - X_i) - n \ln B(1 - \alpha, \alpha).$$

Taking the derivative with respect to α we get

$$\frac{\partial \ln \mathcal{L}(\alpha|X)}{\partial \alpha} = \sum_{i=1}^n \ln \frac{1 - X_i}{X_i} + n\pi \cot(\pi\alpha).$$

Therefore, maximum-likelihood function is

$$\hat{\alpha} = -\frac{1}{\pi} \operatorname{atan} \left(\frac{n\pi}{\sum_{i=1}^n \ln \frac{1-X_i}{X_i}} \right).$$

If $\hat{\alpha}$ is negative, we add 1, because $\frac{\operatorname{atan}}{\pi} \in (-0.5, 0.5)$, while $\alpha \in (0, 1)$.

2.2 Balding-Nichols distribution

Notation:

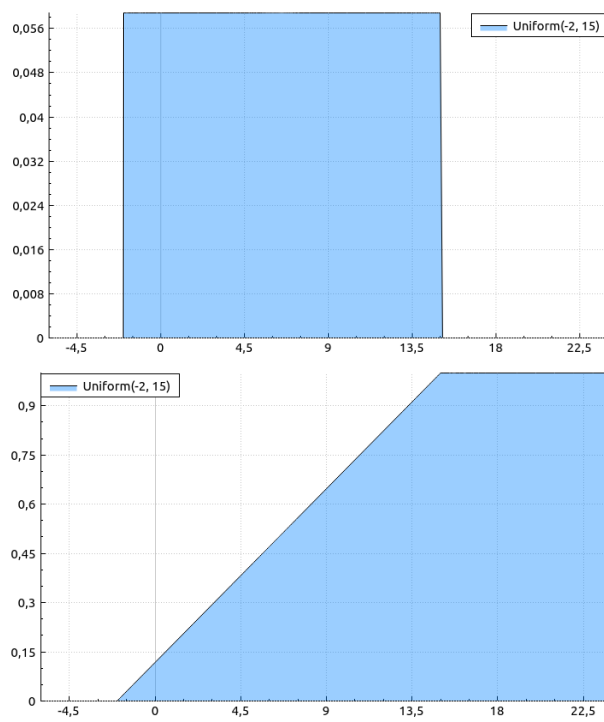
$$X \sim \text{Balding-Nichols}(p, F)$$

with $p, F \in (0, 1)$. Relation to Beta distribution:

$$X \sim \mathcal{B}(pF', (1-p)F')$$

with $F' = (1-F)/F$.

2.3 Uniform distribution



Notation	$X \sim \mathcal{U}(a, b)$
Parameters	$a, b \in \mathbb{R}$
Domain	$x \in [a, b]$
$f(x)$	$\frac{1}{b-a}$
$F(x)$	$\frac{x-a}{b-a}$
$\mathbb{E}[X]$	$\frac{a+b}{2}$
$\text{Var}(X)$	$\frac{(b-a)^2}{12}$
Median	$\frac{a+b}{2}$
Mode	doesn't exist
$\phi(t)$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$

Relation to Beta distribution:

$$X \sim \mathcal{B}(1, 1, a, b).$$

Estimation of support.

Frequentist inference. Likelihood function is

$$\mathcal{L}(a, b|X) = \frac{1}{(b-a)^n} \mathbf{1}_{\{X_i \in [a, b] \ \forall i=1, \dots, n\}}.$$

Therefore, $\mathcal{L}(a, b|X)$ is the largest for $\hat{b} = X_{(n)}$ and $\hat{a} = X_{(1)}$. However, using the fact that $X_{(k)} \sim B(k, n+1-k, a, b)$, these are biased estimators:

$$\mathbb{E}[X_{(1)}] = \frac{an+b}{n+1} \quad \text{and} \quad \mathbb{E}[X_{(n)}] = \frac{a+bn}{n+1}.$$

To get unbiased estimators we make the transformations:

$$\tilde{a} = \frac{nX_{(1)} - X_{(n)}}{n-1} \quad \text{and} \quad \tilde{b} = \frac{nX_{(n)} - X_{(1)}}{n-1}.$$

Then we get

$$\mathbb{E}[\tilde{a}] = \frac{n\mathbb{E}[X_{(1)}] - \mathbb{E}[X_{(n)}]}{n-1} = \frac{n(an+b) - (a+bn)}{n^2-1} = a.$$

Analogously, $\mathbb{E}[\tilde{b}] = b$.

Bayesian inference. Let us say, we try to estimate $\theta = b - a$ with known a . We set the prior distribution $\theta \sim \text{Pareto}(\alpha, \sigma)$:

$$h(\theta|\alpha, \sigma) = \frac{\alpha\sigma^\alpha}{\theta^{\alpha+1}} \mathbf{1}_{\{\theta \geq \sigma\}}.$$

The density of posterior distribution is

$$f(\theta|X) \propto \frac{\alpha\sigma^\alpha}{\theta^{\alpha+n+1}} \mathbf{1}_{\{\theta \geq \max(\sigma, X_{(n)} - a)\}} \sim \text{Pareto}(\alpha + n, \max(\sigma, X_{(n)} - a)).$$

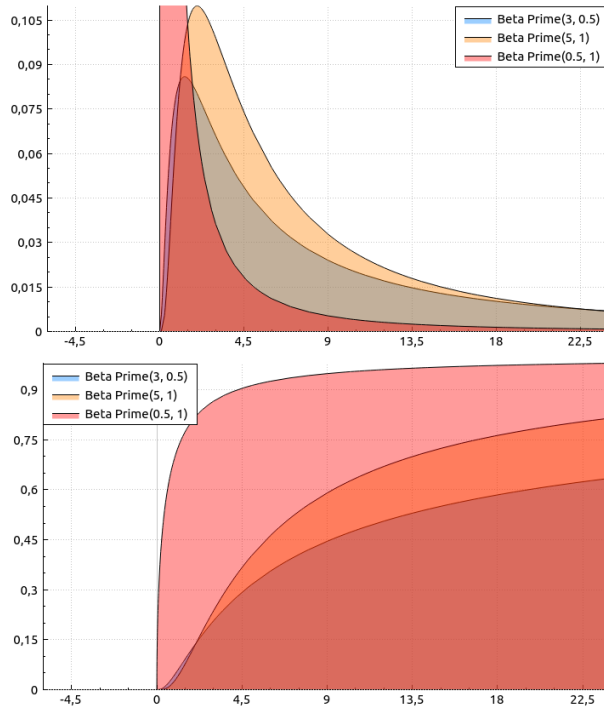
Hence, Bayesian estimator is

$$\mathbb{E}[\theta|X] = \frac{\alpha + n}{\alpha + n - 1} \max(\sigma, X_{(n)} - a)$$

and MAP estimator is

$$\theta_{MAP} = \max(\sigma, X_{(n)} - a).$$

3 Beta-prime distribution



Notation	$X \sim \mathcal{B}'(\alpha, \beta)$
Parameters	$\alpha, \beta > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha, \beta)}$
$F(x)$	$I_{\frac{x}{1+x}}(\alpha, \beta)$
$\mathbb{E}[X]$	$\frac{\alpha}{\beta-1} \mathbf{1}_{\{\beta>1\}} + \infty \mathbf{1}_{\{\beta \leq 1\}}$
$\text{Var}(X)$	$\frac{\alpha(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2}$, if $\beta > 1$
Median	Searched numerically
Mode	$\max\left(\frac{\alpha-1}{\beta+1}, 0\right)$.
$\phi(t)$	Calculated numerically

Relation to other distributions:

$$\frac{X}{1+X} \sim \mathcal{B}(\alpha, \beta),$$

$$\frac{\beta}{\alpha} X \sim F(2\alpha, 2\beta).$$

Search of the median. For $\alpha = \beta$ we have $m = 1$. Otherwise, we use the relation $m = \frac{m'}{1-m'}$, where m' is the median of beta-distribution $\mathcal{B}(\alpha, \beta)$.

Calculation of characteristic function. For $\alpha \geq 1$ one can use numerical integration from section For $\alpha < 1$ we have $\lim_{x \rightarrow 0} f(x) \rightarrow \infty$ and $\int_0^\infty \cos(tx)f(x)dx$ is impossible to compute directly. Then we split the integral:

$$\int_0^\infty \cos(tx)f(x)dx = \int_0^\infty (\cos(tx) - 1)f(x)dx + 1.$$

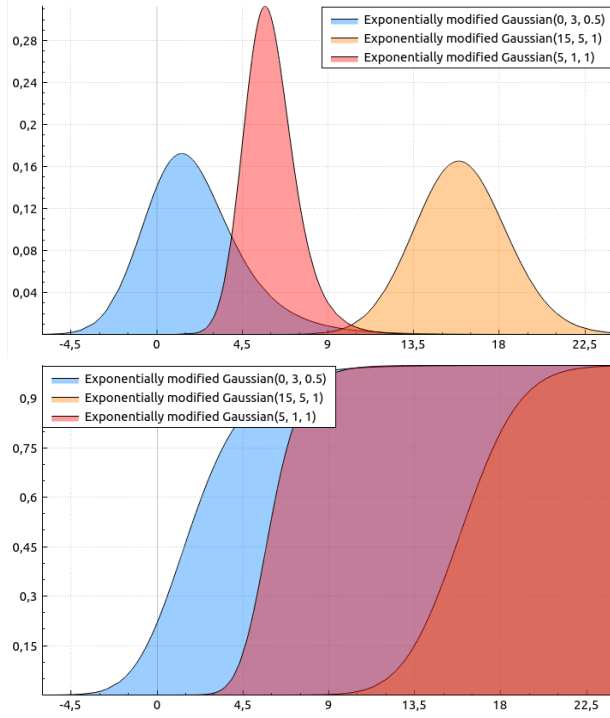
The limit of the integrand for $x \rightarrow 0$ is 0 now, regardless of the value of the shape α .

Estimation of shapes. Using relationship with Beta distribution we transform the sample:

$$Y_i = \frac{X_i}{1 + X_i}, \quad 1 \leq i \leq N,$$

and run estimation for beta-distributed Y .

4 Exponentially-modified Gaussian distribution



Notation	$X \sim \text{EMG}(\mu, \sigma, \lambda)$
Parameters	$\mu \in \mathbb{R}, \sigma > 0, \lambda > 0$
Domain	$x \in \mathbb{R}$
$f(x)$	$\frac{\lambda}{2} e^{\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2x)} \operatorname{erfc}\left(\frac{\mu + \lambda\sigma^2 - x}{\sqrt{2}\sigma^2}\right)$
$F(x)$	$\Phi(u, 0, v) - e^{-u + \frac{v^2}{2} + \log \Phi(u, v^2, v)}$, where $\Phi(x, \mu, \sigma)$ is Gaussian CDF, $u = \lambda(x - \mu)$, $v = \lambda\sigma$.
$\mathbb{E}[X]$	$\mu + 1/\lambda$
$\operatorname{Var}(X)$	$\sigma^2 + 1/\lambda^2$
Median	Searched numerically
Mode	Searched numerically
$\phi(t)$	$\left(1 - \frac{it}{\lambda}\right)^{-1} \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$

Relation to other distribution: if $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \operatorname{Exp}(\lambda)$, then $X + Y \sim \text{EMG}(\mu, \sigma, \lambda)$.

5 F-distribution

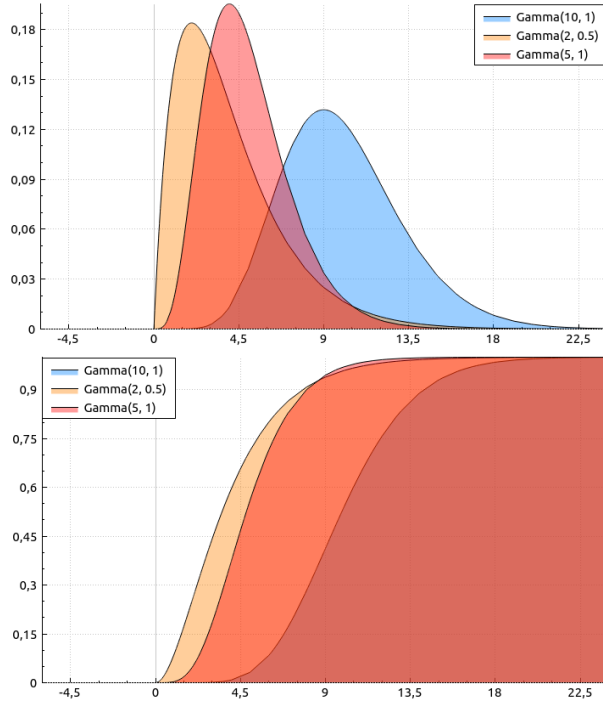
Notation	$X \sim F(d_1, d_2)$
Parameters	$d_1, d_2 > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$
$F(x)$	$I_{\frac{d_1 x}{d_1 x + d_2}}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)$
$\mathbb{E}[X]$	$\frac{d_2}{d_2 - 2}$ for $d_2 > 2$
$\text{Var}(X)$	$\frac{2d_2^2(d_1 + d_2 - 2)}{d_1(d_2 - 2)^2(d_2 - 4)}$ for $d_2 > 4$
Median	Searched numerically
Mode	$\max\left(\frac{d_2(d_1 - 2)}{d_1(d_1 + 2)}, 0\right)$
$\phi(t)$	Calculated numerically

Relation to other distributions:

$$\frac{d_1 X}{d_2 + d_1 X} \sim \mathcal{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right),$$

$$\frac{d_1}{d_2} X \sim \mathcal{B}'\left(\frac{d_1}{2}, \frac{d_2}{2}\right).$$

6 Gamma distribution



Notation	$X \sim \Gamma(\alpha, \beta)$
Parameters	$\alpha > 0, \beta > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
$F(x)$	$P(\alpha, \beta x)$
$\mathbb{E}[X]$	$\frac{\alpha}{\beta}$
$\text{Var}(X)$	$\frac{\alpha}{\beta^2}$
Median	Searched numerically
Mode	$\max\left(\frac{\alpha-1}{\beta}, 0\right)$
$\phi(t)$	$\left(1 - \frac{it}{\beta}\right)^{-\alpha}$

Estimation of parameters.

Frequentist inference. Log-likelihood function:

$$\ln \mathcal{L}(\alpha, \beta | X) = n\alpha \ln \beta - n \ln \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^n \ln X_i - \beta \sum_{i=1}^n X_i.$$

Derivatives:

$$\frac{\partial \ln \mathcal{L}(\alpha, \beta | X)}{\partial \alpha} = n \ln \beta - n\psi(\alpha) + \sum_{i=1}^n \ln X_i,$$

$$\frac{\partial \ln \mathcal{L}(\alpha, \beta | X)}{\partial \beta} = \frac{n\alpha}{\beta} - \sum_{i=1}^n X_i.$$

While the solution for the second equation is analytic:

$$\hat{\beta} = \frac{\alpha}{\bar{X}_n},$$

the first equation is solved numerically, using second derivative:

$$\frac{\partial^2 \ln \mathcal{L}(\alpha, \beta | X)}{\partial \alpha^2} = -n\psi_1(\alpha),$$

or if β is unknown:

$$\frac{\partial^2 \ln \mathcal{L}(\alpha, \beta | X)}{\partial \alpha^2} = -n\psi_1(\alpha) + \frac{n}{\alpha},$$

Moreover, the maximum-likelihood estimation of rate β is biased. Unbiased estimator would be

$$\tilde{\beta} = \frac{\alpha}{\bar{X}_n} \left(1 - \frac{1}{n}\right).$$

Bayesian inference. We suppose that prior distribution of rate β is $\Gamma(\kappa, \gamma)$:

$$h(\beta) = \frac{\gamma^\kappa}{\Gamma(\kappa)} \beta^{\kappa-1} e^{-\gamma\beta}.$$

Then

$$f(\beta | X) \propto \beta^{\alpha n} e^{-\beta \sum_{i=1}^n X_i} \cdot \beta^{\kappa-1} e^{-\gamma\beta} \sim \Gamma\left(\alpha n + \kappa, \gamma + \sum_{i=1}^n X_i\right).$$

Therefore, Bayesian estimator is

$$\mathbb{E}[\beta | X] = \frac{\alpha n + \kappa}{\gamma + \sum_{i=1}^n X_i},$$

and MAP estimator is

$$\beta_{MAP} = \frac{\alpha n + \kappa - 1}{\gamma + \sum_{i=1}^n X_i}.$$

Exponential family parameterization Logarithm of probability mass function:

$$\log \mathbb{P}(X = x) = \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log x - \beta x.$$

Therefore, sufficient statistics $T(x) = (\log x, x)^T$, natural parameters $\theta = (\alpha - 1, -\beta)$, log-normalizer $F(\theta) = \log \Gamma(\theta_1 + 1) - (\theta_1 + 1) \log(-\theta_2)$, carrier measure $k(x) = 0$. Gradient of log-normalizer is $\nabla F(\theta) = (\psi(\theta_1 + 1) - \log(-\theta_2), -\frac{\theta_1 + 1}{\theta_2})^T$. We conclude that adjusted cross-entropy is

$$\begin{aligned} H_F(\theta_q \| \theta_p) &= F(\theta_q) - \langle \theta_q, \nabla F(\theta_p) \rangle \\ &= \log \Gamma(\theta_{q1} + 1) - (\theta_{q1} + 1) \log(-\theta_{q2}) - \theta_{q1} (\psi(\theta_{p1} + 1) - \log(-\theta_{p2})) + \frac{\theta_{q2}(\theta_{p1} + 1)}{\theta_{p2}}. \end{aligned}$$

Adjusted entropy is

$$\begin{aligned} H_F(\theta) &= \log \Gamma(\theta_1 + 1) - \log(-\theta_2) - \theta_1 \psi(\theta_1 + 1) + \theta_1 + 1 \\ &= \log \Gamma(\alpha) - \log \beta - (\alpha - 1) \cdot \psi(\alpha) + \alpha. \end{aligned}$$

And Kullback-Leibler divergence:

$$\begin{aligned}\text{KL}(p\|q) &= H_F(\theta_q\|\theta_p) - H_F(\theta_p) \\ &= \log \frac{\Gamma(\alpha_q)}{\Gamma(\alpha_p)} + \alpha_q \log \frac{\beta_p}{\beta_q} + (\alpha_p - \alpha_q)\psi(\alpha_p) + \alpha_p \left(\frac{\beta_q}{\beta_p} - 1 \right)\end{aligned}$$

6.1 Chi-squared distribution

Notation:

$$X \sim \chi_k^2.$$

Relation to Gamma distribution:

$$X \sim \Gamma\left(\frac{k}{2}, \frac{1}{2}\right).$$

Kullback-Leibler divergence:

$$\text{KL}(p\|q) = \log \frac{\Gamma(k_q/2)}{\Gamma(k_p/2)} + \frac{1}{2}(k_p - k_q)\psi(k_p/2).$$

Relation to other distributions: if $X_1, \dots, X_k \sim \mathcal{N}(0, 1)$, then $\sum_{i=1}^k X_i^2 \sim \chi_k^2$.

6.2 Erlang distribution

Notation:

$$X \sim \text{Erlang}(k, \beta).$$

The only difference between Gamma and Erlang distributions is that latter takes an integer number k as a shape parameter.

6.3 Exponential distribution

Relation to Gamma distribution:

$$X \sim \Gamma(1, \lambda).$$

Hence, estimation of parameter λ is the particular case of estimation of rate β for Gamma distribution.

Adjusted cross-entropy:

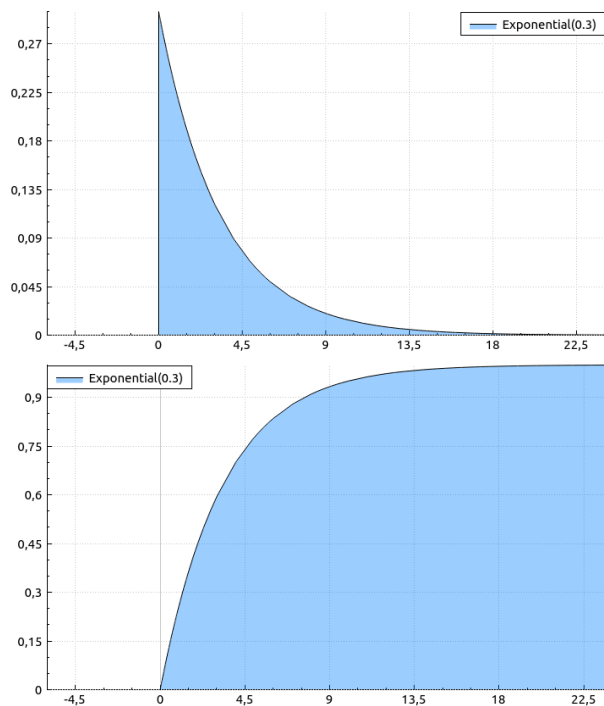
$$H_F(\lambda_q\|\lambda_p) = \log \lambda_q + \frac{\lambda_q}{\lambda_p}.$$

Thus adjusted entropy is

$$H_F(\lambda) = \log \lambda$$

and Kullback-Leibler divergence:

$$\text{KL}(p\|q) = \log \frac{\lambda_p}{\lambda_q} + \left(\frac{\lambda_q}{\lambda_p} - 1 \right).$$

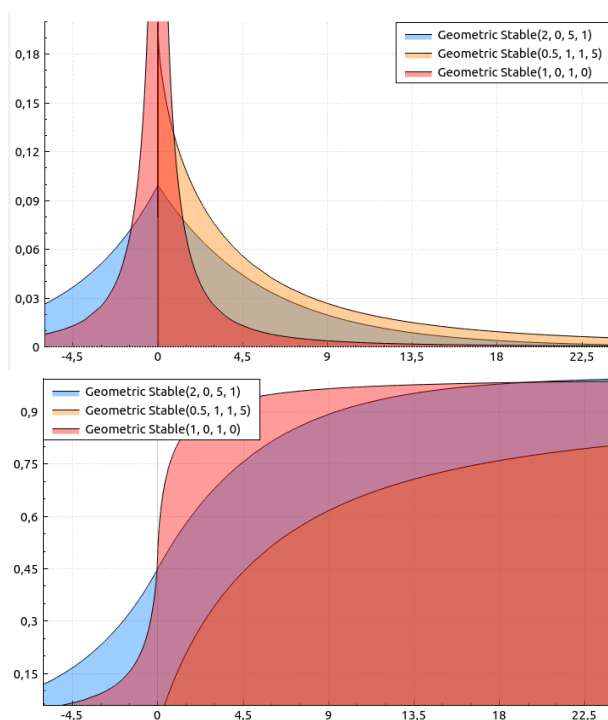


Notation	$X \sim \text{Exp}(\lambda)$
Parameters	$\lambda > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\lambda e^{-\lambda x}$
$F(x)$	$1 - e^{-\lambda x}$
$\mathbb{E}[X]$	$\frac{1}{\lambda}$
$\text{Var}(X)$	$\frac{1}{\lambda^2}$
Median	$\frac{\ln(2)}{\lambda}$
Mode	0
$\phi(t)$	$\frac{\lambda}{\lambda - it}$

7 Geometric Stable distribution

7.1 Asymmetric Laplace distribution

7.2 Laplace distribution



Notation	$X \sim \text{GS}_\alpha(\beta, \gamma, \mu)$
Parameters	$\alpha \in (0, 2], \beta \in [-1, 1],$ $\gamma > 0, \mu \in \mathbb{R}$
Domain	$x \in \dots$
$f(x)$	Calculated numerically
$F(x)$	Calculated numerically
$\mathbb{E}[X]$	$k + \lambda$
$\text{Var}(X)$	$2(k + 2\lambda)$
Median	Searched numerically
Mode	Searched numerically
$\phi(t)$...

8 Kolmogorov-Smirnov distribution

9 Logistic distribution

10 Log-normal distribution

11 Marchenko-Pastur distribution

Notation	$X \sim \mathcal{MP}(\lambda, \sigma^2)$
Parameters	$\lambda, \sigma^2 > 0$
Domain	$x \in [\sigma^2 a, \sigma^2 b]$, if $\lambda < 1$, $x \in [\sigma^2 a, \sigma^2 b] \cup \{0\}$, otherwise, where $a = (1 - \sqrt{\lambda})^2$ and $b = (1 + \sqrt{\lambda})^2$
$f(x)$...
$F(x)$...
$\mathbb{E}[X]$	σ^2
$\text{Var}(X)$	$\sigma^4 \lambda$
Median	0 if $\lambda > 2$, otherwise searched numerically
Mode	$\frac{\sigma^2(\lambda-1)^2}{\lambda+1}$, if $\lambda < 1$, 0, otherwise
$\phi(t)$	Calculated numerically

Calculation of characteristic function. For $\lambda > 1$ we use numerical integration by definition

$$\phi(t) = \int_{\sigma^2 a}^{\sigma^2 b} \cos(tx) f(x) dx + i \int_{\sigma^2 a}^{\sigma^2 b} \sin(tx) f(x) dx.$$

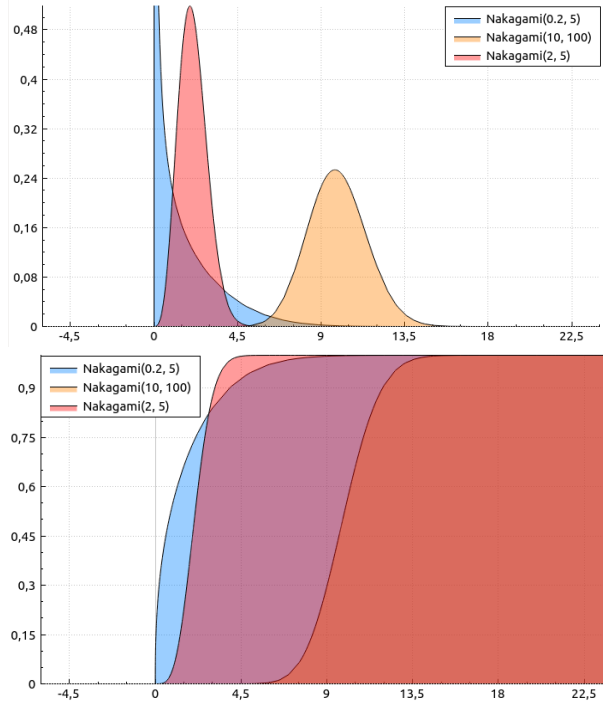
For $\lambda = 1$ we split the integrand for real part by $(\cos(tx) - 1)f(x)$ and $f(x)$:

$$\Re(\phi(t)) = \int_{\sigma^2 a}^{\sigma^2 b} (\cos(tx) - 1) f(x) dx + 1.$$

And for $\lambda < 1$ we calculate integral at point 0 separately:

$$\begin{aligned} \phi(t) &= \int_{\{0\} \cup [\sigma^2 a, \sigma^2 b]} \cos(tx) f(x) dx + i \int_{\{0\} \cup [\sigma^2 a, \sigma^2 b]} \sin(tx) f(x) dx \\ &= 1 - \frac{1}{\lambda} + \int_{\sigma^2 a}^{\sigma^2 b} \cos(tx) f(x) dx + i \int_{\sigma^2 a}^{\sigma^2 b} \sin(tx) f(x) dx. \end{aligned}$$

12 Nakagami distribution



Notation	$X \sim \text{Nakagami}(\mu, \omega)$
Parameters	$\mu, \omega > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\frac{2\mu^\mu}{\Gamma(\mu)\omega^\mu} x^{2\mu-1} e^{-\frac{\mu}{\omega}x^2}$
$F(x)$	$P(\mu, \mu x^2 / \omega)$
$\mathbb{E}[X]$	$\frac{\Gamma(\mu+1/2)}{\Gamma(\mu)} \sqrt{\frac{\omega}{\mu}}$
$\text{Var}(X)$	$\omega - (\mathbb{E}[X])^2$
Median	Searched numerically
Mode	$\max\left(\sqrt{\frac{(2\mu-1)\omega}{2\mu}}, 0\right)$
$\phi(t)$	Calculated numerically

Calculation of characteristic function. For $\mu < 1$ $\lim_{x \rightarrow 0} f(x) \rightarrow \infty$. Then we use the following transformation for real part of characteristic function:

$$\begin{aligned} \Re(\phi(t)) &= \int_0^\infty \cos(tx) f(x) dx \\ &= \int_0^\infty (\cos(tx) - 1) f(x) dx + 1 \end{aligned}$$

Relation to other distributions: if $Y \sim \Gamma(\mu, \mu/\omega)$, then

$$X \sim \text{Nakagami}(\mu, \omega).$$

12.1 Chi distribution

Notation:

$$X \sim \chi_k$$

Relation to Nakagami distribution:

$$X \sim \text{Nakagami}(k/2, k).$$

12.2 Maxwell-Boltzmann distribution

Notation:

$$X \sim \text{MB}(\sigma)$$

Relation to Nakagami distribution:

$$X \sim \text{Nakagami}(3/2, \sigma^2).$$

12.3 Rayleigh distribution

Notation:

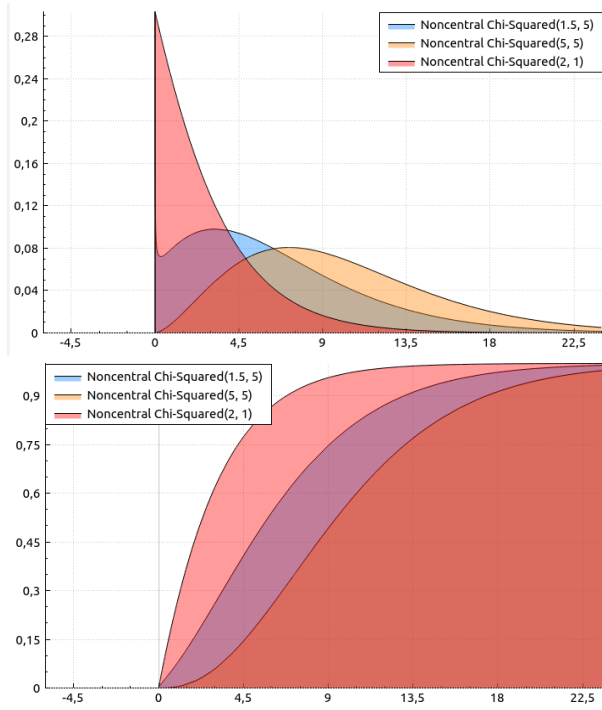
$$X \sim \text{Rayleigh}(\sigma)$$

Relation to Nakagami distribution:

$$X \sim \text{Nakagami}(1, 2\sigma^2).$$

Estimation of scale. ...

13 Noncentral Chi-Squared distribution



Notation	$X \sim \chi_k'^2(\lambda)$
Parameters	$k > 0, \lambda > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\frac{1}{2} e^{-\frac{x+\lambda}{2}} \left(\frac{x}{\lambda}\right)^{\frac{k-2}{4}} I_{\frac{k}{2}-1}(\sqrt{\lambda x})$
$F(x)$	$\text{MarcumP}_{\frac{k}{2}}\left(\frac{\lambda}{2}, \frac{x}{2}\right)$
$\mathbb{E}[X]$	$k + \lambda$
$\text{Var}(X)$	$2(k + 2\lambda)$
Median	Searched numerically
Mode	Searched numerically for $k > 2$, 0, otherwise
$\phi(t)$	$\frac{\exp \frac{-it\lambda}{1-2it}}{(1-2it)^{k/2}}$

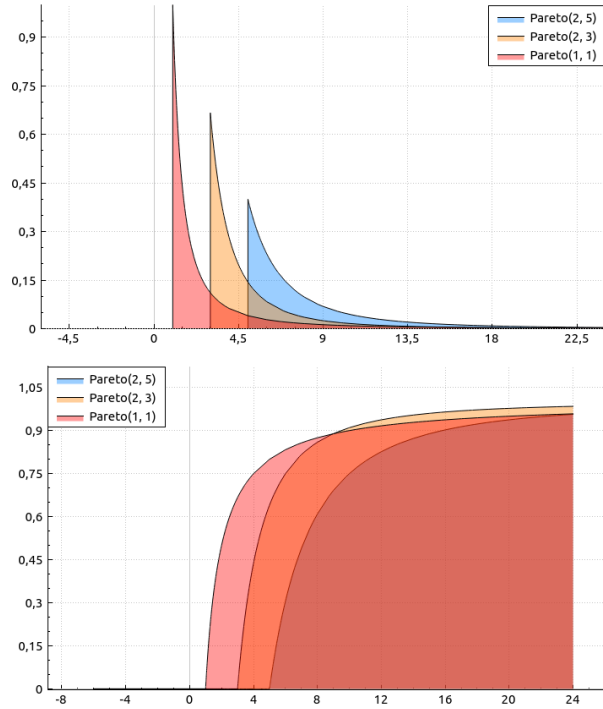
Relation to other distributions:

- Let X_1, \dots, X_k be independent with $X_i \sim \mathcal{N}(\mu_i, 1)$, $i = 1, \dots, k$. Then

$$\sum_{i=1}^k X_i^2 \sim \chi_k'^2\left(\sum_{i=1}^k \mu_i^2\right).$$

- If $\lambda = 0$, then $X \sim \chi_k^2$.
- If $J \sim \text{Po}(\lambda)$, then $\chi_{k+2J}^2 \sim \chi_k'^2(\lambda)$.

14 Pareto distribution



Notation	$X \sim \text{Pareto}(\alpha, \sigma)$
Parameters	$\alpha, \sigma > 0$
Domain	$x \geq \sigma$
$f(x)$	$\frac{\alpha \sigma^\alpha}{x^{\alpha+1}}$
$F(x)$	$1 - \left(\frac{\sigma}{x}\right)^\alpha$
$\mathbb{E}[X]$	$\frac{\alpha \sigma}{\alpha - 1}$ for $\alpha > 1$, ∞ otherwise
$\text{Var}(X)$	$\frac{\sigma^2 \alpha}{(\alpha - 1)^2 (\alpha - 2)}$ for $\alpha > 2$, ∞ otherwise
Median	$\sigma 2^{1/\alpha}$
Mode	σ
$\phi(t)$	Calculated numerically

Estimation of parameters.

Frequentist inference. Log-likelihood function is

$$\ln \mathcal{L}(\alpha, \sigma | X) = n \ln \alpha + n \alpha \ln \sigma - (\alpha + 1) \sum_{i=1}^n \ln X_i.$$

We assume that $\sigma \leq X_{(1)}$, otherwise sample X couldn't have been generated from such distribution. It is obvious, that $\ln \mathcal{L}(\alpha, \sigma | X)$ is an increasing function in terms of σ , therefore $\hat{\sigma} = X_{(1)}$ is an optimal estimator. Let's take derivative with respect to α :

$$\frac{\partial \ln \mathcal{L}(\alpha, \sigma | X)}{\partial \alpha} = \frac{n}{\alpha} + n \ln \sigma - \sum_{i=1}^n \ln X_i.$$

From this we conclude that the maximum-likelihood estimator of shape is

$$\hat{\alpha} = \frac{1}{\frac{1}{n} \left(\sum_{i=1}^n \ln X_i \right) - \ln \hat{\sigma}}.$$

It is known that $\hat{\sigma} \sim \text{Pareto}(n\alpha, \sigma)$ and $\hat{\alpha} \sim \text{Inv-}\Gamma(n-1, n\alpha)$ and they are independent. Then

$$\mathbb{E}[\hat{\sigma}] = \frac{\sigma}{1 - \frac{1}{n\alpha}}$$

and

$$\mathbb{E}[\hat{\alpha}] = \frac{n\alpha}{n-2}.$$

Therefore, in order to get unbiased estimators we need to make the following transformations:

$$\tilde{\alpha} = \frac{n-2}{n}\hat{\alpha} \quad \text{and} \quad \tilde{\sigma} = \hat{\sigma} \left(1 - \frac{1}{(n-1)\hat{\alpha}}\right).$$

Note that if we estimate parameters separately, then $\hat{\alpha} \sim \text{Inv-}\Gamma(n, n\alpha)$ and transformations are different.

Bayesian inference. We now assume that σ is known and prior distribution of α is $\Gamma(\kappa, \beta)$:

$$h(\alpha) = \frac{\beta^\kappa}{\Gamma(\kappa)} \alpha^{\kappa-1} e^{-\beta\alpha}.$$

The density of posterior distribution is

$$f(\alpha|X) \propto \prod_{i=1}^n \frac{\sigma^\alpha}{X_i^{\alpha-1}} \cdot \alpha^{\kappa+n-1} e^{-\beta\alpha} \propto \alpha^{\kappa+n-1} e^{-(\beta + \sum_{i=1}^n \ln(X_i/\sigma))\alpha}.$$

Therefore, $\alpha|X \sim \Gamma(\kappa + n, \beta + \sum_{i=1}^n \ln(X_i/\sigma))$ and Bayesian estimator is

$$\mathbb{E}[\alpha|X] = \frac{\kappa + n}{\beta + \sum_{i=1}^n \ln(X_i/\sigma)}.$$

MAP estimator is

$$\alpha_{MAP} = \frac{\kappa + n - 1}{\beta + \sum_{i=1}^n \ln(X_i/\sigma)}.$$

Note on fitting scale with Bayes: let it be vice versa, α is known while σ is not. Then we say that a priori $\sigma \sim \text{Pareto}(\kappa, \theta)$:

$$h(\sigma) = \frac{\kappa\theta^\kappa}{\sigma^{\kappa+1}}.$$

Then posterior distribution is:

$$f(\sigma|X) \propto \prod_{i=1}^n \frac{1}{X_i^{\alpha-1}} \cdot \sigma^{\alpha n - \kappa - 1} \mathbf{1}_{\{\theta < \sigma < X_{(1)}\}} \sim \text{Bounded-Pareto}(\kappa - \alpha n, \theta, X_{(1)}).$$

This imposes the following additional constraints on the prior hyperparameters: $\kappa > \alpha n$ and $\theta < X_{(1)}$. Bayesian estimator:

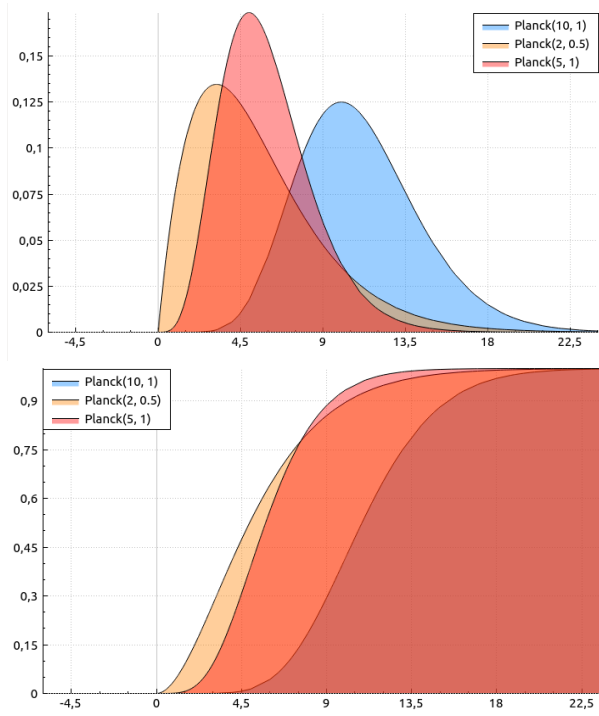
$$\mathbb{E}[\sigma|X] = \frac{\theta^{\alpha'}}{1 - \left(\frac{\theta}{X_{(1)}}\right)^{\alpha'}} \cdot \left(\frac{\alpha'}{\alpha' - 1}\right) \cdot \left(\frac{1}{\theta^{\alpha'}} - \frac{1}{X_{(1)}^{\alpha'}}\right)$$

with $\alpha' = \kappa - \alpha n$. MAP estimator is just

$$\sigma_{MAP} = \theta.$$

However, Bounded-Pareto distribution is not yet supported in RandLib.

15 Planck distribution



Notation	$X \sim \text{Planck}(a, b)$
Parameters	$a, b > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\frac{b^{a+1}}{\Gamma(a+1)\zeta(a+1)} \cdot \frac{x^a}{e^{bx}-1}$
$F(x)$	Calculated numerically
$\mathbb{E}[X]$	$\frac{(a+1)\zeta(a+2)}{b\zeta(a+1)}$
$\text{Var}(X)$	$\frac{(a+1)(a+2)\zeta(a+3)}{b^2\zeta(a+1)} - (\mathbb{E}[X])^2$
Median	Searched numerically
Mode	$\frac{W_0(-ae^{-a})+a}{b}$, if $a > 1$, otherwise 0
$\phi(t)$	Calculated numerically

Calculation of cumulative distribution function. For $a \geq 1$ $F(x)$ can be calculated by straightforward numerical integration:

$$F(x) = \frac{b^{a+1}}{\Gamma(a+1)\zeta(a+1)} \int_0^x \frac{t^a}{e^{bt}-1} dt.$$

Note that for $a < 1$ integrand has a singularity point at $t = 0$. In such case we define

$$h(t) = \frac{b^{a+2}t^{a+1}}{\Gamma(a+1)\zeta(a+1)} \cdot \left(\frac{1}{e^{bt}-1} - \frac{1}{bt} \right)$$

and then

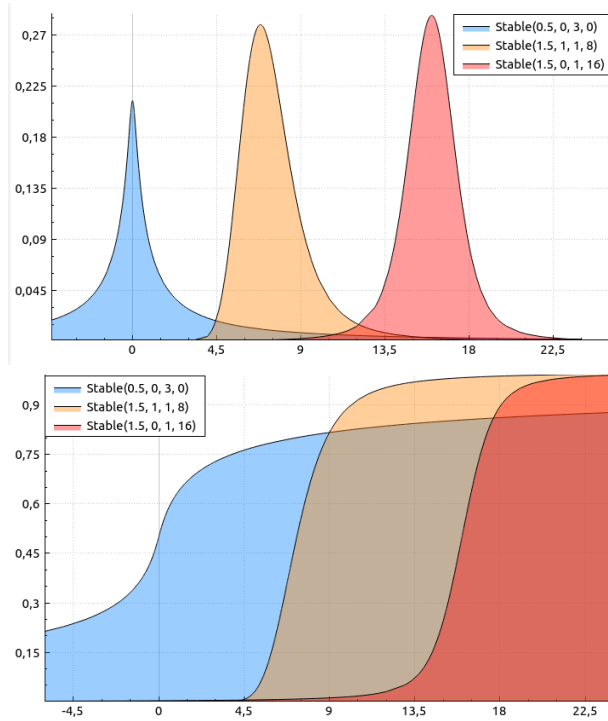
$$F(x) = \int_0^x h(t)dt + \frac{(bx)^a}{a\Gamma(a+1)\zeta(a+1)}.$$

Calculation of characteristic function. The idea of calculations for $a < 1$ is near the same. We split the real part of $\phi(t)$ into 3 different integrals:

$$\Re(\phi(t)) = \int_0^1 \cos(tx)h(x)dx + \int_1^\infty \cos(tx)f(x)dx + \frac{b^a}{a\Gamma(a+1)\zeta(a+1)} \left(\cos(t) + t \int_0^1 \sin(tx)x^a dx \right).$$

All the integrands now have no singularity points.

16 Stable distribution



Notation	$X \sim S_\alpha(\beta, \gamma, \mu)$
Parameters	$\alpha \in (0, 2], \beta \in [-1, 1],$ $\gamma > 0, \mu \in \mathbb{R}$
Domain	$x \in \mathbb{R}$, if $\beta \neq 1$, $x \in [\mu, \infty)$, if $\beta = 1, \alpha < 2$, $x \in (-\infty, \mu]$, if $\beta = -1, \alpha < 2$
$f(x)$	Calculated numerically
$F(x)$	Calculated numerically
$\mathbb{E}[X]$	μ for $\alpha > 1$, otherwise undefined
$\text{Var}(X)$	$2\gamma^2 1_{\{\alpha=2\}} + \infty 1_{\{\alpha<2\}}$
Median	μ for $\beta = 0$, otherwise searched numerically
Mode	μ , if $\beta = 0$ or $\alpha = 2$, $\mu + \frac{\beta\gamma}{3}$, if $ \beta = 1$ and $\alpha = \frac{1}{2}$, otherwise searched numerically
$\phi(t)$...

Calculation of p.d.f.

Calculation of c.d.f.

16.1 Cauchy distribution

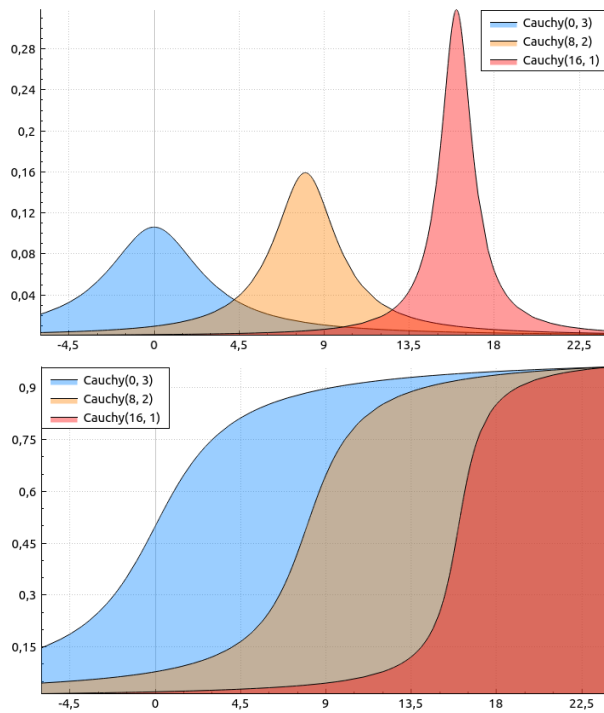
Relation to Stable distribution:

$$X \sim S_1(0, \gamma, \mu)$$

16.2 Levy distribution

Relation to Stable distribution:

$$X \sim S_{\frac{1}{2}}(1, \gamma, \mu)$$



Notation	$X \sim \text{Cauchy}(\mu, \gamma)$
Parameters	$\mu \in \mathbb{R}, \gamma^2 > 0$
Domain	$x \in \mathbb{R}$
$f(x)$	$\frac{1}{\pi\gamma \left[1 + \left(\frac{x-\mu}{\gamma} \right)^2 \right]}$
$F(x)$	$\frac{1}{\pi} \text{atan} \left(\frac{x-\mu}{\gamma} \right) + \frac{1}{2}$
$\mathbb{E}[X]$	Undefined
$\text{Var}(X)$	∞
Median	μ
Mode	μ
$\phi(t)$	$e^{i\mu t - \gamma t }$

16.3 Normal distribution

Relation to Stable distribution:

$$X \sim S_2(\cdot, \sigma^2/2, \mu)$$

Estimation of parameters

Frequentist inference. Maximum-likelihood estimators for Normal distribution are very well-known:

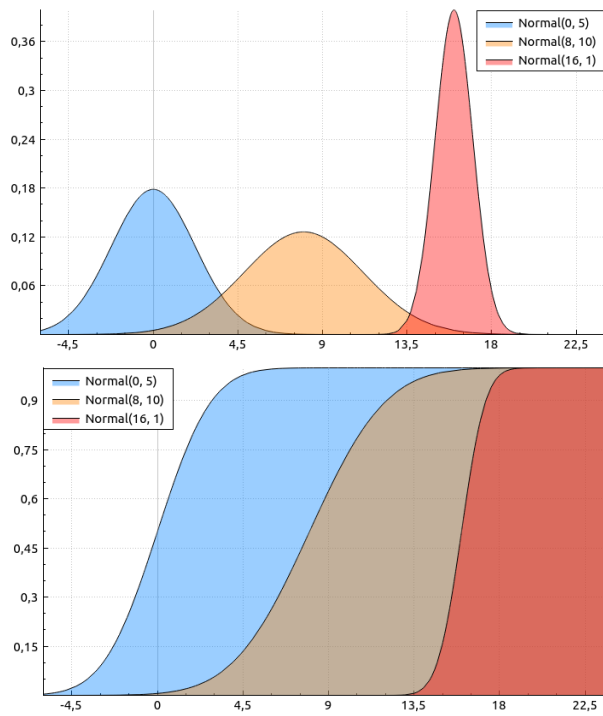
$$\hat{\mu} = \overline{X}_n \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

However, for unknown μ the value of $\hat{\sigma}^2 \sim \frac{\sigma^2}{n} \chi_{n-1}^2$. Therefore, unbiased estimator in this case would be

$$\widetilde{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

Moreover, if one is interested in estimating scale σ with known μ , then maximum likelihood estimator is

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \sim \frac{\sigma}{\sqrt{n}} \chi_n$$



Notation	$X \sim \mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}, \sigma^2 > 0$
Domain	$x \in \mathbb{R}$
$f(x)$	$\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
$F(x)$	$\frac{1}{2} \operatorname{erfc}\left(\frac{\mu-x}{\sqrt{2\sigma^2}}\right)$
$\mathbb{E}[X]$	μ
$\operatorname{Var}(X)$	σ^2
Median	μ
Mode	μ
$\phi(t)$	$e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$

and

$$\mathbb{E}[\hat{\sigma}] = \frac{\sigma}{\sqrt{n}} \sqrt{2} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}.$$

Then unbiased estimator is

$$\tilde{\sigma} = \hat{\sigma} \sqrt{\frac{n}{2}} \frac{\Gamma(n/2)}{\Gamma((n+1)/2)}$$

Bayesian inference. ...

16.4 Holtsmark distribution

Relation to Stable distribution:

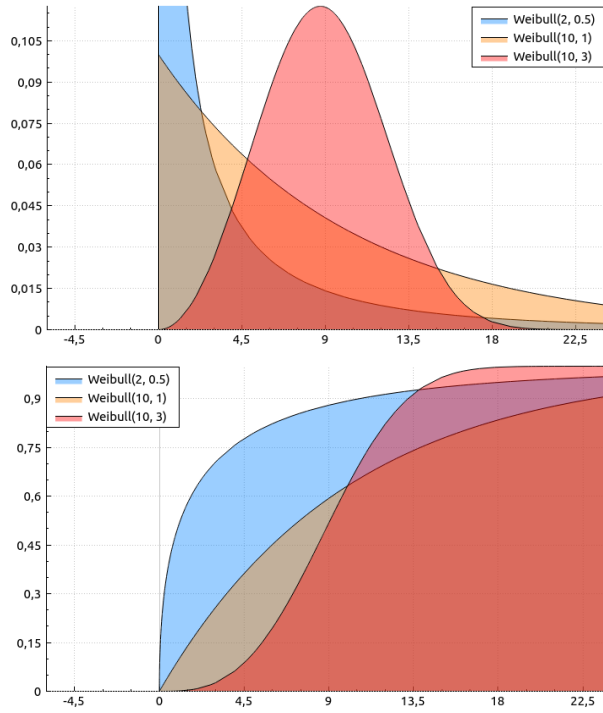
$$X \sim S_{\frac{3}{2}}(0, \gamma, \mu)$$

16.5 Landau distribution

Relation to Stable distribution:

$$X \sim S_1(1, \gamma, \mu)$$

17 Weibull



Notation	$X \sim \text{Weibull}(\lambda, k)$
Parameters	$\lambda, k > 0$
Domain	$x \in \mathbb{R}^+$
$f(x)$	$\frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp(-(x/\lambda)^k)$
$F(x)$	$1 - \exp(-(x/\lambda)^k)$
$\mathbb{E}[X]$	$\lambda \Gamma(1 + 1/k)$
$\text{Var}(X)$	$\lambda^2 \Gamma(1 + 2/k) - (\mathbb{E}[X])^2$
Median	$\lambda (\ln 2)^{\frac{1}{k}}$
Mode	$\lambda \left(1 - \frac{1}{k}\right)^{\frac{1}{k}}$
$\phi(t)$	Calculated numerically

Estimation of scale

Frequentist inference. Log-likelihood function:

$$\ln \mathcal{L}(\lambda, k | X) = n(\ln k - \ln \lambda) + (k-1) \sum_{i=1}^n (\ln X_i - \ln \lambda) - \frac{1}{\lambda^k} \sum_{i=1}^n X_i^k.$$

The derivative with respect to scale:

$$\frac{\partial \ln \mathcal{L}(\lambda, k | X)}{\partial \lambda} = -\frac{nk}{\lambda} + \frac{k}{\lambda^{k+1}} \sum_{i=1}^n X_i^k = 0.$$

Therefore, maximum-likelihood estimation for λ is

$$\hat{\lambda} = \left(\frac{1}{n} \sum_{i=1}^n X_i^k \right)^{\frac{1}{k}}.$$

Bayesian inference. Assume k is known. Instead of estimating λ we give an estimation for λ^k . Let's say that prior distribution of λ^k is $\text{Inv-}\Gamma(\alpha, \beta)$:

$$h(\lambda^k) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-k(\alpha+1)} e^{-\beta/\lambda^k}.$$

Posterior distribution then:

$$f(\lambda^k|X) \propto \lambda^{-k(\alpha+1+n)} e^{-\frac{1}{\lambda^k}(\beta + \sum_{i=1}^n X_i^k)} \sim \text{Inv-}\Gamma\left(\alpha + n, \beta + \sum_{i=1}^n X_i^k\right).$$

Bayesian estimator:

$$\mathbb{E}[\lambda^k|X] = \frac{\beta + \sum_{i=1}^n X_i^k}{\alpha + n - 1},$$

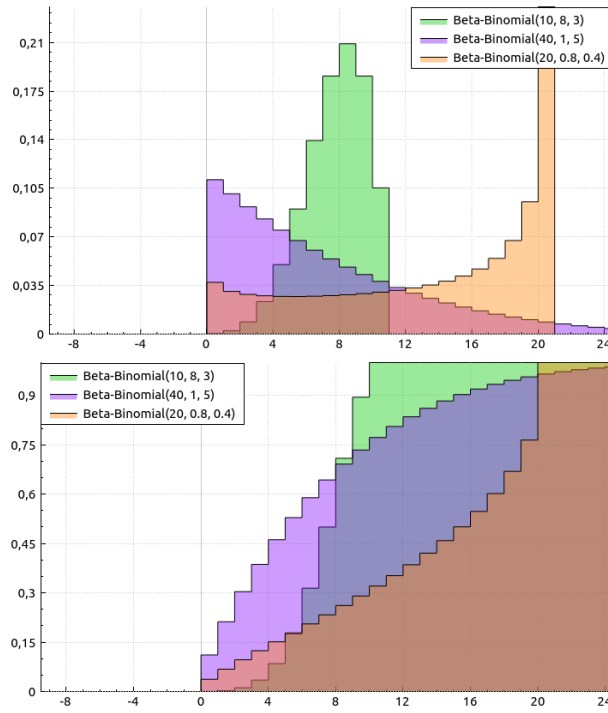
MAP estimator:

$$\lambda_{MAP}^k = \frac{\beta + \sum_{i=1}^n X_i^k}{\alpha + n + 1}.$$

Part III

Discrete univariate distributions

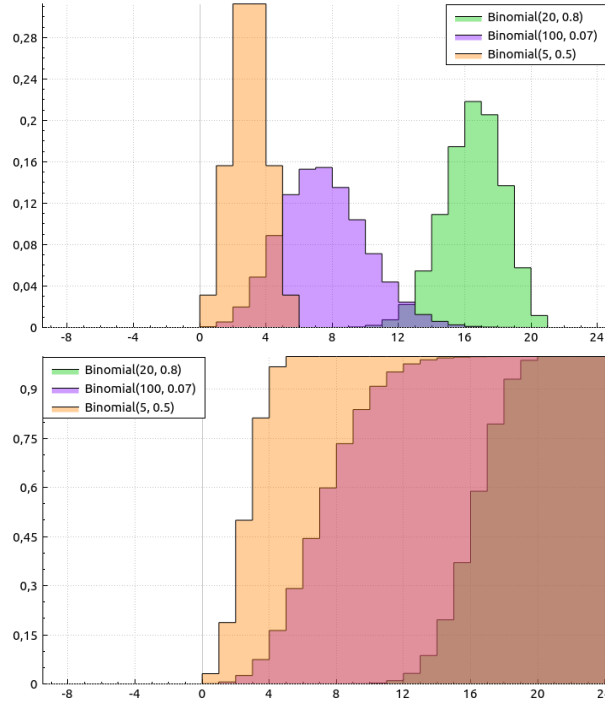
18 Beta-binomial distribution



Notation	$X \sim \text{BB}(n, \alpha, \beta)$
Parameters	$n \in \mathbb{N}, \alpha, \beta > 0$
Domain	$k \in \{0, \dots, n\}$
$\mathbb{P}(X = k)$	$\binom{n}{k} \frac{B(k+\alpha, n-k+\beta)}{B(\alpha, \beta)}$
$\mathbb{P}(X \leq k)$	Calculated numerically
$\mathbb{E}[X]$	$n \frac{\alpha}{\alpha + \beta}$
$\text{Var}(X)$	$\frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Median	Searched numerically
Mode	Searched numerically
$\phi(t)$	Calculated numerically

Relation to other distributions: if $p \sim \mathcal{B}(\alpha, \beta)$, then $\text{Bin}(n, p) \sim \text{BB}(n, \alpha, \beta)$.

19 Binomial distribution



Notation	$X \sim \text{Bin}(n, p)$
Parameters	$n \in \mathbb{N}, p \in [0, 1]$
Domain	$k \in \{0, \dots, n\}$
$\mathbb{P}(X = k)$	$\binom{n}{k} p^k (1 - p)^{n-k}$
$\mathbb{P}(X \leq k)$	$I_{1-p}(n - k, 1 + k)$
$\mathbb{E}[X]$	np
$\text{Var}(X)$	$np(1 - p)$
Median	$[np]$
Mode	$[(n + 1)p]$
$\phi(t)$	$(1 - p + pe^{it})^n$

Estimation of probability p with known number n .

Frequentist inference. Log-likelihood function:

$$\ln \mathcal{L}(p|X) \propto \sum_{i=1}^k (X_i \log p + (n - X_i) \log(1 - p))$$

The derivative with respect to p is:

$$\frac{\partial \ln \mathcal{L}(p|X)}{\partial p} = \frac{\sum_{i=1}^k X_i}{p} - \frac{nk - \sum_{i=1}^k X_i}{1 - p}.$$

Therefore we reach the maximum value of log-likelihood if

$$p = \frac{\bar{X}_k}{n}.$$

Bayesian inference. We set prior Beta distribution $\mathcal{B}(\alpha, \beta)$:

$$h(p) = \frac{p^{\alpha-1} (1 - p)^{\beta-1}}{B(\alpha, \beta)}.$$

Then posterior is

$$f(p|X) \propto p^{\alpha-1+\sum_{i=1}^k X_i} (1-p)^{\beta-1+\sum_{i=1}^k (n-X_i)} \sim \mathcal{B}\left(\alpha + \sum_{i=1}^k X_i, \beta + nk - \sum_{i=1}^k X_i\right).$$

Thus Bayesian estimator is

$$\mathbb{E}[p|X] = \frac{\alpha + \sum_{i=1}^k X_i}{\alpha + \beta + nk}$$

and MAP estimator is

$$p_{MAP} = \frac{\alpha + \sum_{i=1}^k X_i - 1}{\alpha + \beta + nk - 2}.$$

Also, Minimax estimator is equal to Bayes estimator if $\alpha = \beta = \frac{1}{2}\sqrt{n}$.

19.1 Bernoulli

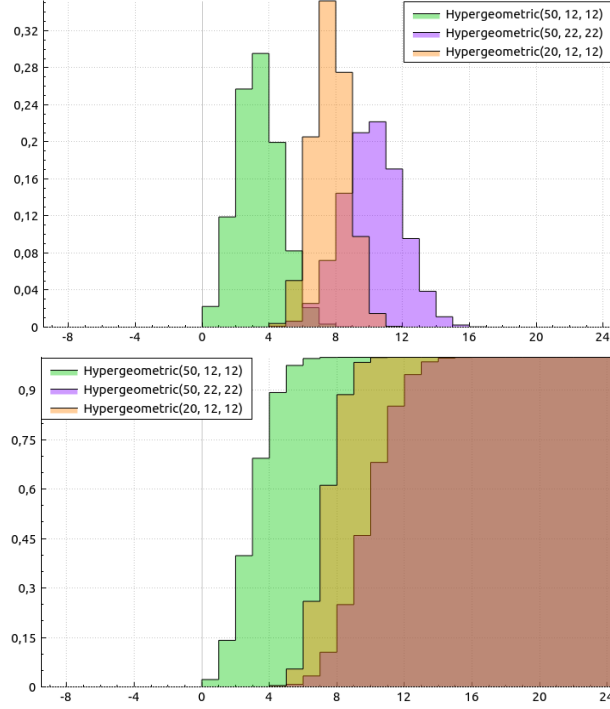
Notation:

$$X \sim \text{Bernoulli}(p).$$

Relation to Binomial distribution:

$$X \sim \text{Bin}(1, p).$$

20 Hypergeometric distribution



Notation	$X \sim \text{HG}(N, K, n)$
Parameters	$N \in \mathbb{N}, K \in \{1, 2, \dots, N\},$ $n \in \{1, 2, \dots, N\}$
Domain	$\max(0, n + K - N) \leq k \leq \min(n, K)$
$\mathbb{P}(X = k)$	$\frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$
$\mathbb{P}(X \leq k)$	Calculated numerically
$\mathbb{E}[X]$	$\frac{nK}{N}$
$\text{Var}(X)$	$\frac{nK(N-K)(N-n)}{N^2(N-1)}$
Median	Searched numerically
Mode	$\left\lfloor \frac{(n+1)(K+1)}{N+2} \right\rfloor$
$\phi(t)$	Calculated numerically

Estimation of number of target members of population K .

Bayesian inference. Let prior distribution of K be Beta-Binomial distribution $BB(N, \alpha, \beta)$:

$$h(K) = \binom{N}{K} \frac{B(K + \alpha, N - K + \beta)}{B(\alpha, \beta)}.$$

Then for one sample X :

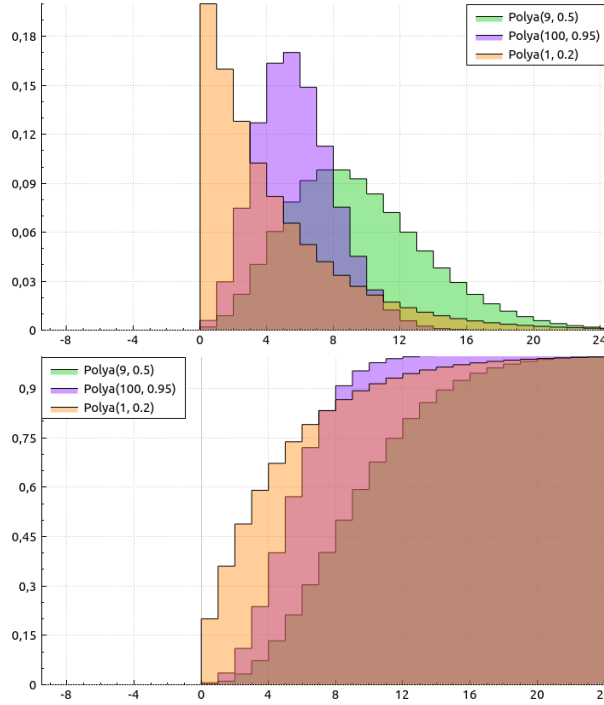
$$K - X \sim BB(N - n, \alpha + X, \beta + nk - X)$$

and therefore

$$\mathbb{E}[K|X] = X + (N - n) \frac{\alpha}{\alpha + \beta}.$$

However, RandLib doesn't support Bayesian fitting for Hypergeometric distribution yet.

21 Negative-Binomial (Polya) distribution



Notation	$X \sim \text{NB}(r, p)$
Parameters	$r > 0, p \in (0, 1)$
Domain	$k \in \mathbb{N}_0$
$\mathbb{P}(X = k)$	$\binom{k+r-1}{k} p^r (1-p)^k$
$\mathbb{P}(X \leq k)$	$I_p(r, k+1)$
$\mathbb{E}[X]$	$\frac{1-p}{p} r$
$\text{Var}(X)$	$\frac{1-p}{p^2} r$
Median	Searched numerically
Mode	$\max \left(\left\lfloor \frac{(r-1)(1-p)}{p} \right\rfloor, 0 \right)$
$\phi(t)$	$\left(\frac{p}{1-(1-p)e^{it}} \right)^r$

Relation to other distributions: if $\lambda \sim \text{Gamma} \left(r, \frac{p}{1-p} \right)$, then $\text{Po}(\lambda) \sim \text{NB}(r, p)$.

21.1 Geometric distribution

Notation:

$$X \sim \text{Geometric}(p).$$

Relation to Negative-Binomial distribution:

$$X \sim \text{NB}(1, p).$$

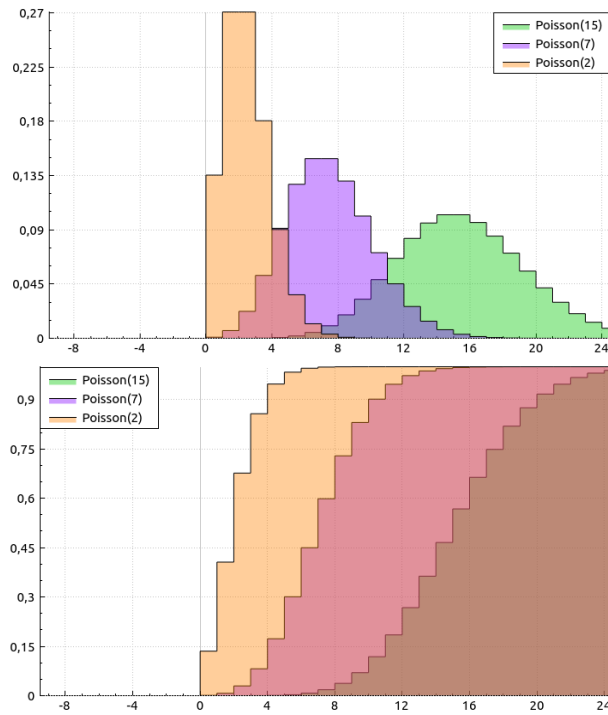
21.2 Pascal distribution

Notation:

$$X \sim \text{Pascal}(r, p).$$

The only difference with Negative-Binomial distribution is that for Pascal distribution shape r is an integer.

22 Poisson distribution



Notation	$X \sim \text{Po}(\lambda)$
Parameters	$\lambda > 0$
Domain	$k \in \mathbb{N}_0$
$\mathbb{P}(X = k)$	$\frac{\lambda^k e^{-\lambda}}{k!}$
$\mathbb{P}(X \leq k)$	$Q(k + 1, \lambda)$
$\mathbb{E}[X]$	λ
$\text{Var}(X)$	λ
Median	$\sim \max\left(\left[\lambda + \frac{1}{3} - \frac{0.02}{\lambda}\right], 0\right)$
Mode	$[\lambda]$
$\phi(t)$	$\exp\{\lambda(e^{it} - 1)\}$

Estimation of rate.

Frequentist inference. Log-likelihood function:

$$\ln \mathcal{L}(\lambda|X) \propto -\lambda n + \sum_{i=1}^n X_i \log \lambda.$$

Setting the derivative w.r.t. rate to 0 we get the optimal value:

$$\lambda = \overline{X}_n.$$

Bayesian inference. Let set prior distribution of $\lambda \sim \Gamma(\alpha, \beta)$:

$$h(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}.$$

Posterior distribution:

$$f(\lambda|X) \propto e^{-\lambda(\beta+n)} \lambda^{\alpha-1+\sum_{i=1}^n X_i} \sim \Gamma\left(\alpha + \sum_{i=1}^n X_i, \beta + n\right).$$

Therefore, Bayesian estimator:

$$\mathbb{E}[\lambda|X] = \frac{\alpha + \sum_{i=1}^n X_i}{\beta + n}.$$

And MAP estimator:

$$\lambda_{MAP} = \max\left(\frac{\alpha + \sum_{i=1}^n X_i - 1}{\beta + n}, 0\right).$$

Exponential family parameterization Logarithm of probability mass function:

$$\log \mathbb{P}(X = x) = x \log \lambda - \lambda - \log(x!).$$

Therefore, sufficient statistics $T(x) = x$, natural parameter $\theta = \log \lambda$, log-normalizer $F(\theta) = \exp(\theta)$, carrier measure $k(x) = \log(x!)$. We conclude that adjusted cross-entropy is

$$\begin{aligned} H_F(\theta_q || \theta_p) &= F(\theta_q) - \langle \theta_q, \nabla F(\theta_p) \rangle \\ &= \exp(\theta_q) - \theta_q \exp(\theta_p). \end{aligned}$$

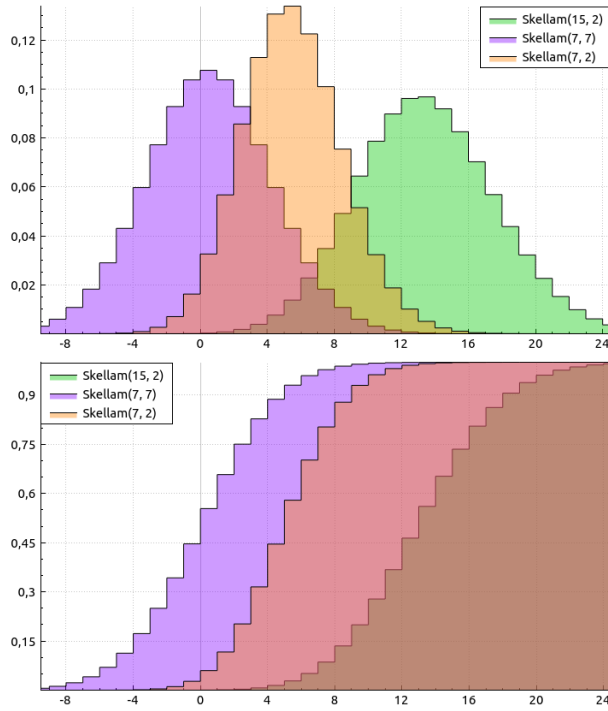
Adjusted entropy is

$$H_F(\theta) = \exp(\theta)(1 - \theta) = \lambda(1 - \log \lambda).$$

And Kullback-Leibler divergence:

$$\begin{aligned} \text{KL}(p||q) &= H_F(\theta_q || \theta_p) - H_F(\theta_p) \\ &= \lambda_q + \lambda_p \left(1 + \log \left(\frac{\lambda_p}{\lambda_q}\right)\right). \end{aligned}$$

23 Skellam distribution

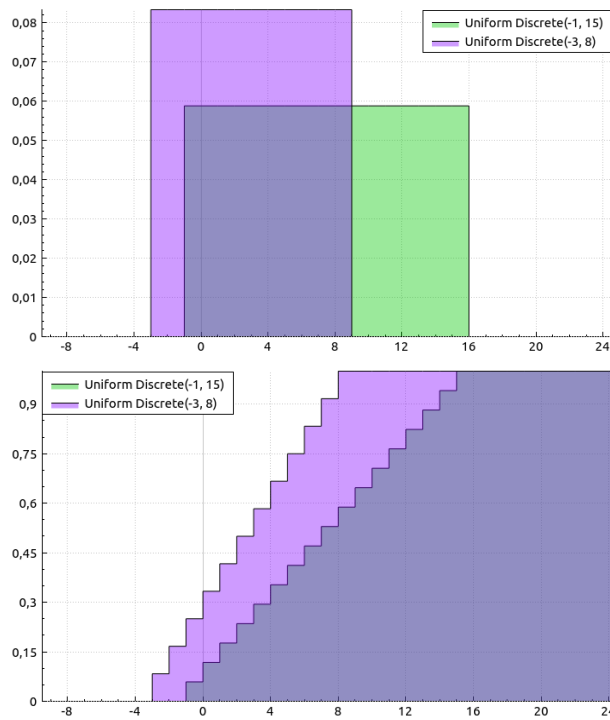


Notation	$X \sim \text{Skellam}(\mu_1, \mu_2)$
Parameters	$\mu_1, \mu_2 > 0$
Domain	$k \in \mathbb{Z}$
$\mathbb{P}(X = k)$	$e^{-(\mu_1 + \mu_2)} \left(\frac{\mu_1}{\mu_2}\right)^{\frac{k}{2}} I_k(2\sqrt{\mu_1 \mu_2})$
$\mathbb{P}(X \leq k)$	$\text{MarcumP}_{k+1}(\mu_2, \mu_1), k \geq 0$ $\text{MarcumQ}_{-k}(\mu_1, \mu_2), k < 0$
$\mathbb{E}[X]$	$\mu_1 - \mu_2$
$\text{Var}(X)$	$\mu_1 + \mu_2$
Median	Searched numerically
Mode	$[\mu_1 - \mu_2]$
$\phi(t)$	$\exp\{\mu_1(e^{it} - 1) - \mu_2(e^{it} - 1)\}$

Relation to other distributions: if $Y \sim \text{Po}(\mu_1)$ and $Z \sim \text{Po}(\mu_2)$, then

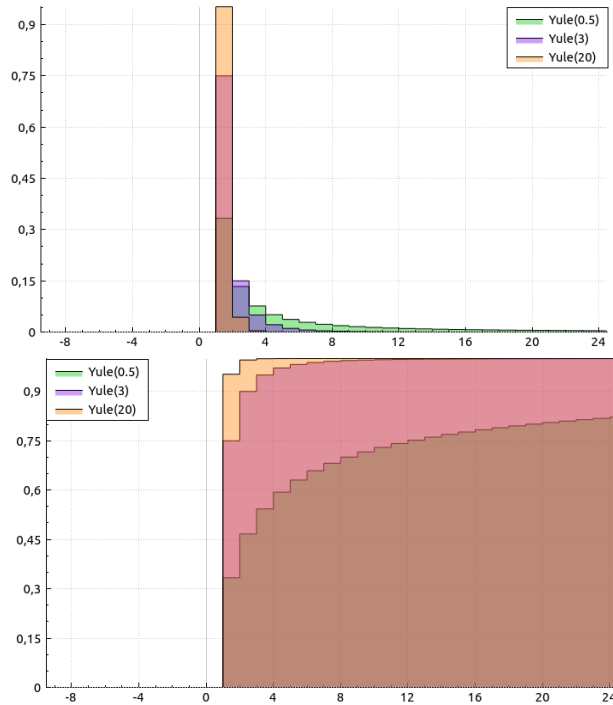
$$Y - Z \sim \text{Skellam}(\mu_1, \mu_2).$$

24 Uniform discrete distribution



Notation	$X \sim U\{a, \dots, b\}$
Parameters	$a, b \in \mathbb{R}, a \leq b$
Domain	$k \in \{a, \dots, b\}$
$\mathbb{P}(X = k)$	$\frac{1}{n}$, where $n = b - a + 1$.
$\mathbb{P}(X \leq k)$	$\frac{k-a+1}{n}$
$\mathbb{E}[X]$	$\frac{a+b}{2}$
$\text{Var}(X)$	$\frac{(n-1)(n+1)}{12}$
Median	$\frac{a+b}{2}$
Mode	Any value between a and b
$\phi(t)$	$\frac{\cos(at) - \cos((b+1)t) + i(\sin(at) - \sin((b+1)t))}{n(1 - \cos(t) - i \sin(t))}$

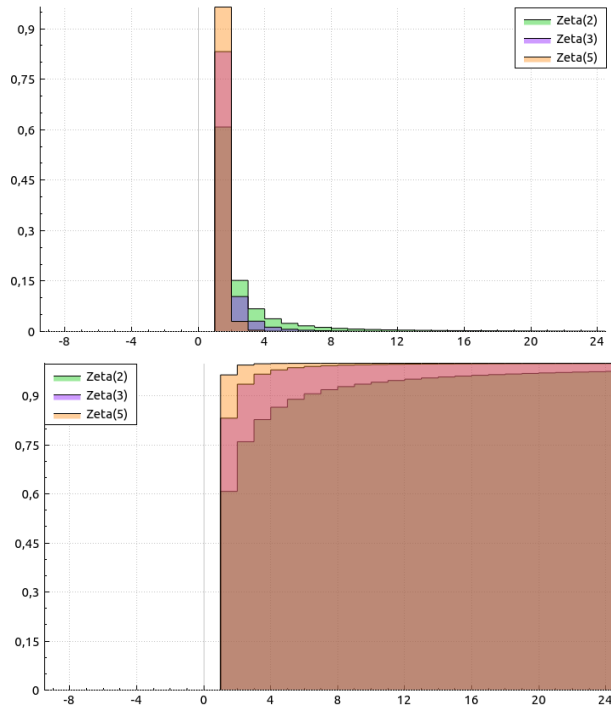
25 Yule distribution



Notation	$X \sim \text{Yule}(\rho)$
Parameters	$\rho \in \mathbb{R}^+$
Domain	$k \in \mathbb{N}$
$\mathbb{P}(X = k)$	$\rho \frac{\Gamma(1+\rho)(k-1)!}{\Gamma(k+\rho+1)}$
$\mathbb{P}(X \leq k)$	$1 - k \frac{\Gamma(1+\rho)(k-1)!}{\Gamma(k+\rho+1)}$
$\mathbb{E}[X]$	$\frac{\rho}{\rho-1}, \rho > 1$ ∞ , otherwise
$\text{Var}(X)$	$\frac{\rho^2}{(\rho-1)^2(\rho-2)}, \rho > 2$ ∞ , otherwise
Median	Searched numerically
Mode	1
$\phi(t)$	Calculated numerically

Relation to other distributions: if $X \sim \text{Pareto}(\alpha, 1)$, then $\text{Geometric}(1/X) \sim \text{Yule}(\alpha)$.

26 Zeta distribution



Notation	$X \sim \text{Zeta}(s)$
Parameters	$s > 1$
Domain	$k \in \mathbb{N}$
$\mathbb{P}(X = k)$	$\frac{1}{\zeta(s)k^s}$
$\mathbb{P}(X \leq k)$	$\frac{H(s,k)}{\zeta(s)}$
$\mathbb{E}[X]$	$\frac{\zeta(s-1)}{\zeta(s)}, s > 2$ ∞ , otherwise
$\text{Var}(X)$	$\frac{\zeta(s-2)}{\zeta(s)} - (\mathbb{E}[X])^2, \rho > 3$ ∞ , otherwise
Median	Searched numerically
Mode	1
$\phi(t)$	Calculated numerically

27 Zipf distribution

Part IV

Bivariate distributions

28 Bivariate Normal distribution

29 Normal-Inverse-Gamma distribution

30 Trinomial distribution

Part V

Circular distributions

31 von Mises distribution

32 Wrapped Exponential distribution

Part VI

Singular distributions

33 Cantor distribution