RandLib documentation

Aleksandr Samarin

November 12, 2017

Contents

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

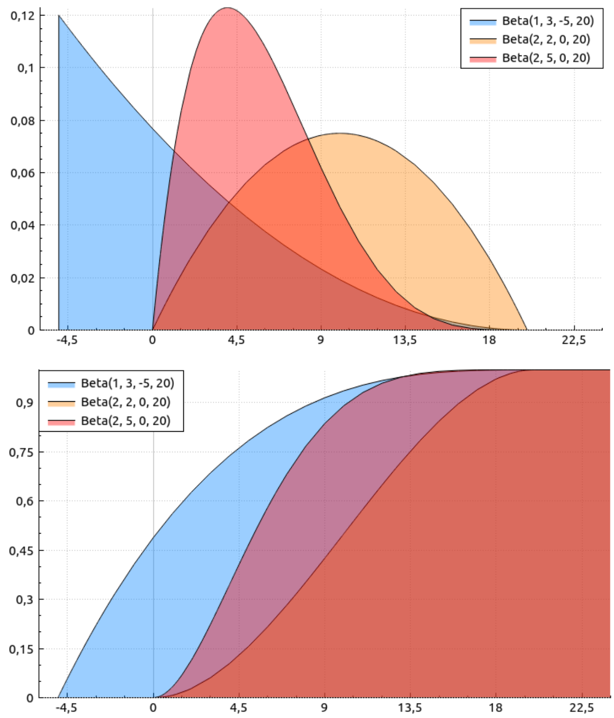
$$\begin{aligned}
\delta &= x - m_1, \\
m_1' &= m_1 + \frac{\delta}{n}, \\
m_2' &= m_2 + \delta^2 \frac{n-1}{n}, \\
m_3' &= m_3 + \delta^3 \frac{(n-1)(n-2)}{n^2} - 3\delta \frac{m_2}{n}, \\
m_4' &= m_4 + \delta^4 \frac{(n-1)(n^2-3n+3)}{n^3} + 6\delta^2 \frac{m_2}{n^2} - 4\delta \frac{m_3}{n}.
\end{aligned}$$

Then m_1', $\frac{m_2}{n}$, $\mathrm{Skew}(X) = \frac{\sqrt{n} m_3'}{m_2'^{3/2}}$ and $\mathrm{Kurt}(X) = \frac{n m_4'}{m_2'^2}$ (we return excess kurtosis).

Part II

Continuous univariate distributions

2 Beta distribution



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

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

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

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

Differentiating again we get the Hessian matrix:

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

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

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

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

These values are applicable only if $\hat{s}_n^2 < \overline{X}_n(1-\overline{X}_n)$. 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

Relation to Beta distribution:

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

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

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

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

2.2 Balding-Nichols distribution

Notation:

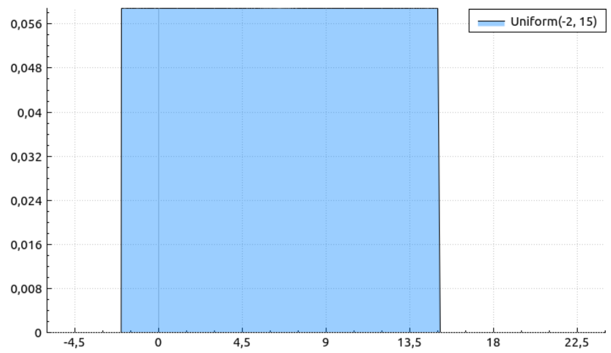
$$X \sim \mathrm{Balding-Nichols}(p, F)$$

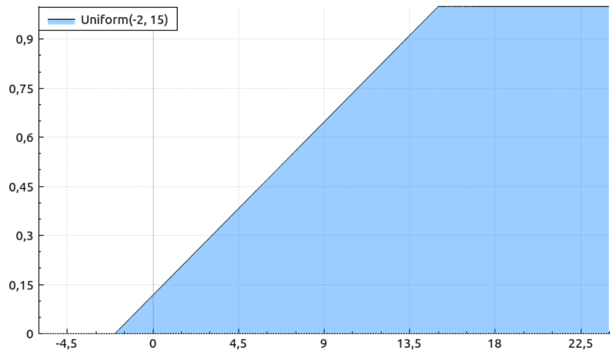
with $p, F \in (0, 1)$. Relation to Beta distribution:

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

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

2.3 Uniform distribution





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

Relation to Beta distribution:

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

Estimation of support.

Frequentist inference. Likelihood function is

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

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

Relation to other distributions:

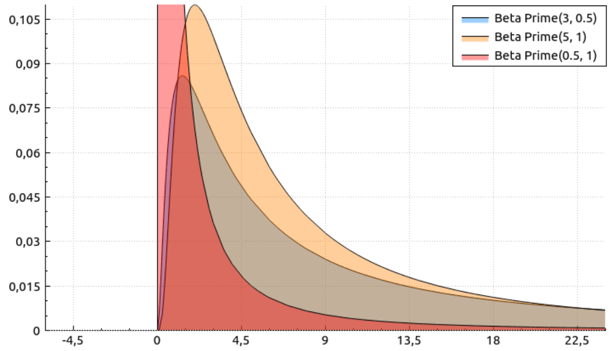
$$\frac{X}{1+X} \sim B(\alpha, \beta),$$

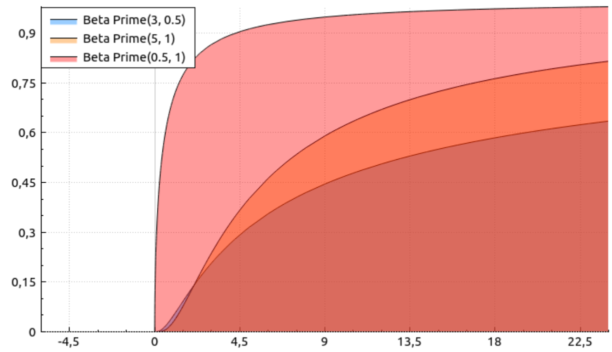
$$\frac{\beta}{\alpha} X \sim F(2\alpha, 2\beta).$$

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 BetaRand estimation for Y.





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

4 Exponentially-modified Gaussian distribution

Notation	$X \sim \mathrm{EMG}(\mu, \sigma, \lambda)$
Parameters	$\mu \in \mathbb{R}, \sigma > 0, \lambda > 0$
Support	$x \in \mathbb{R}$
$f(x)$	...
$F(x)$	...
$\mathbb{E}[X]$	$\mu + 1/\lambda$
$\mathrm{Var}(X)$	$\sigma^2 + 1/\lambda^2$
Median	Searched numerically
Mode	Searched numerically
$\phi(t)$	...

5 F-distribution

Notation	$X \sim \mathrm{F}(d_1, d_2)$
Parameters	$d_1, d_2 > 0$
Support	$x \in \mathbb{R}^+$
$f(x)$	$\frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$
$F(x)$	$I_{\frac{d_1 x}{d_1 x + d_2}}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)$
$\mathbb{E}[X]$	$\frac{d_2}{d_2-2}$ for $d_2 > 2$
$\mathrm{Var}(X)$	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ for $d_2 > 4$
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:

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

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

6 Gamma distribution

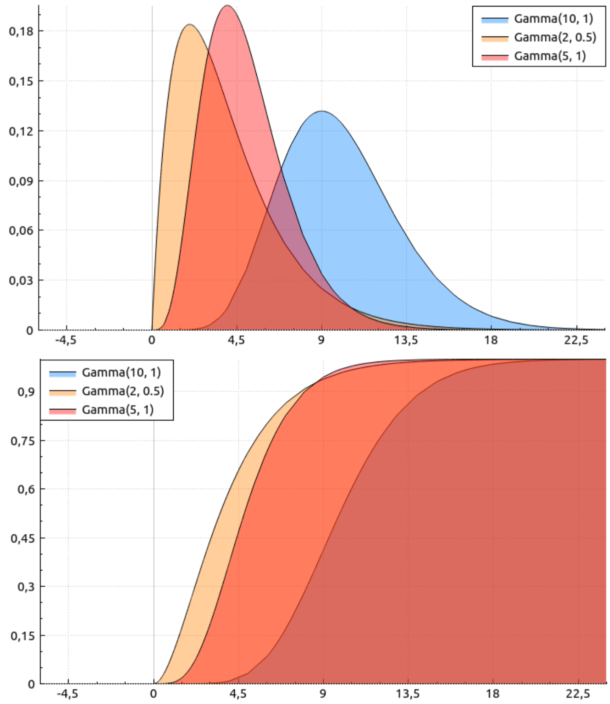
Estimation of parameters.

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:

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



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

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

While the solution for the second equation is analytic:

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

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

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

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

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

Then

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

Therefore, Bayesian estimator is

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

and MAP estimator is

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

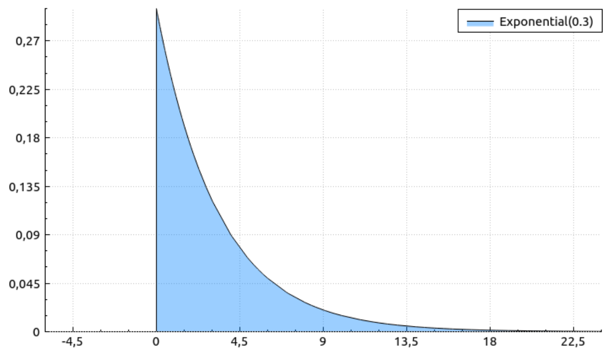
6.1 Chi-squared distribution

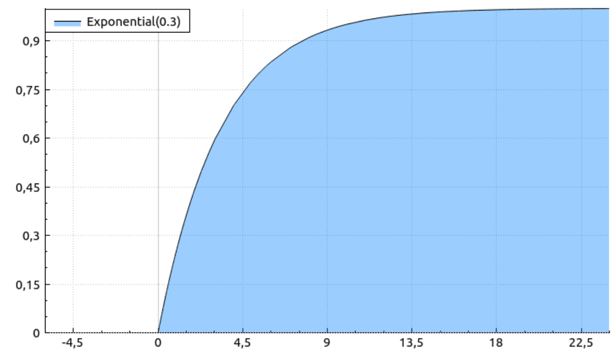
Relation to Gamma distribution:

6.2 Erlang distribution

The only difference between Gamma and Erlang distributions is that a second one takes an integer shape parameter k.

6.3 Exponential distribution

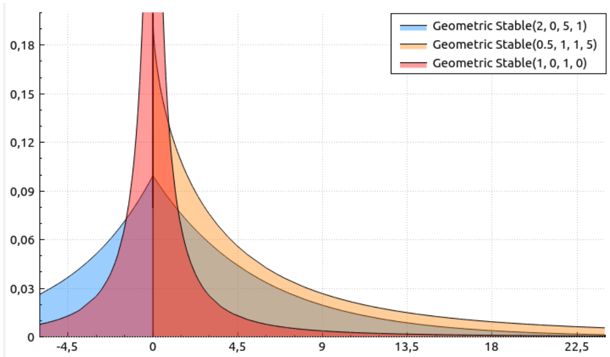


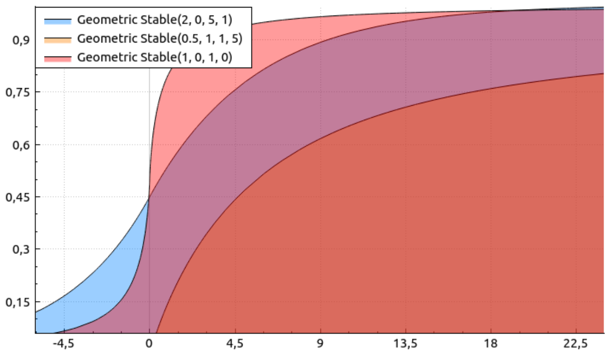


Notation	$X \sim \text{Exp}(\lambda)$
Parameters	$\lambda > 0$
Support	$x \in \mathbb{R}^+$
$f(x)$	$\lambda e^{-\lambda x}$
$F(x)$	$1 - e^{-\lambda x}$
$\mathbb{E}[X]$	$\frac{1}{\lambda}$
$\text{Var}(X)$	$\frac{1}{\lambda^2}$
Median	$\frac{\ln(2)}{\lambda}$
Mode	0
$\phi(t)$	$\frac{\lambda}{\lambda -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

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.

7 Geometric Stable distribution





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

7.1 Asymmetric Laplace distribution

7.2 Laplace distribution

8 Noncentral Chi-Squared distribution

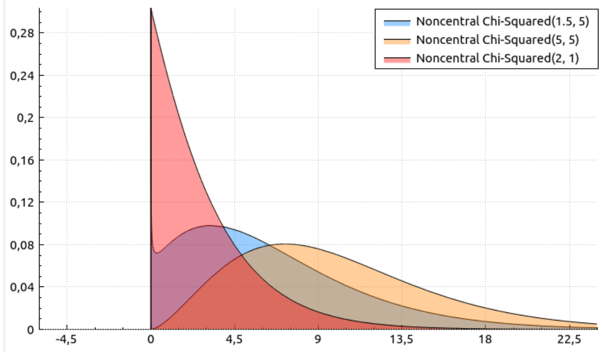
9 Planck distribution

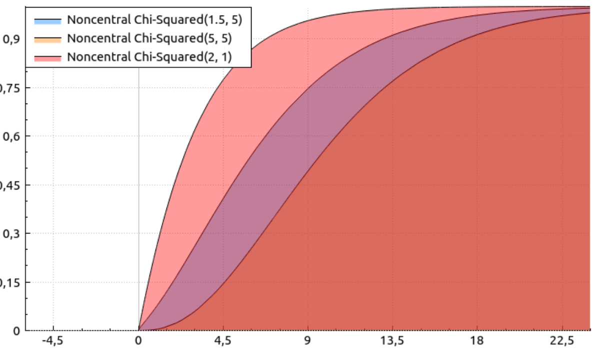
10 Stable distribution

10.1 Normal distribution

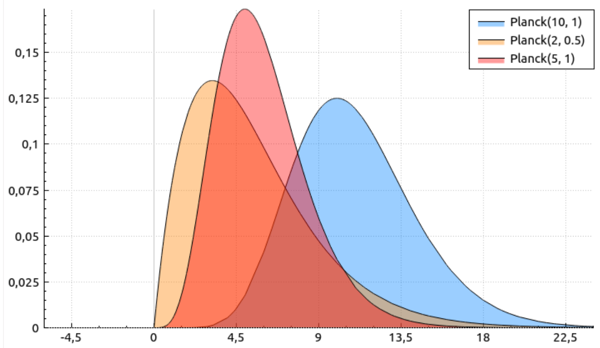
Relation to Stable distribution:

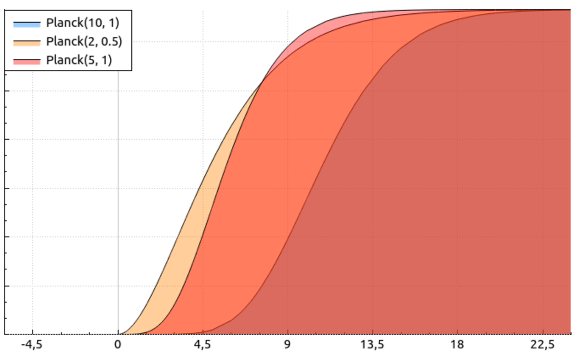
$$X \sim S_2(\cdot, \sigma^2/2, \mu)$$



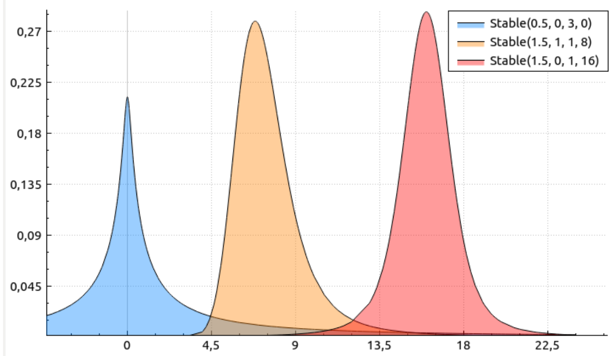


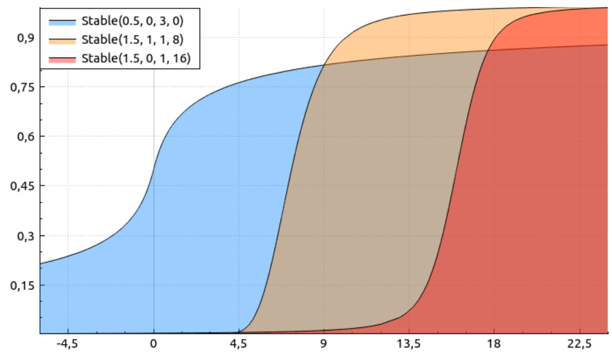
Notation	$X \sim \chi'^2_k(\lambda)$
Parameters	$k > 0, \lambda > 0$
Support	$x \in \mathbb{R}^+$
$f(x)$	$\ldots$
$F(x)$	$P_{\frac{k}{2}}(\ldots)$
$\mathbb{E}[X]$	$k + \lambda$
$\mathrm{Var}(X)$	$2(k + 2\lambda)$
Median	Searched numerically
Mode	Searched numerically
$\phi(t)$	$\frac{\exp \frac{it\lambda}{1-2it}}{(1-2it)^{k/2}}$





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





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

10.2 Cauchy distribution

Relation to Stable distribution:

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

10.3 Levy distribution

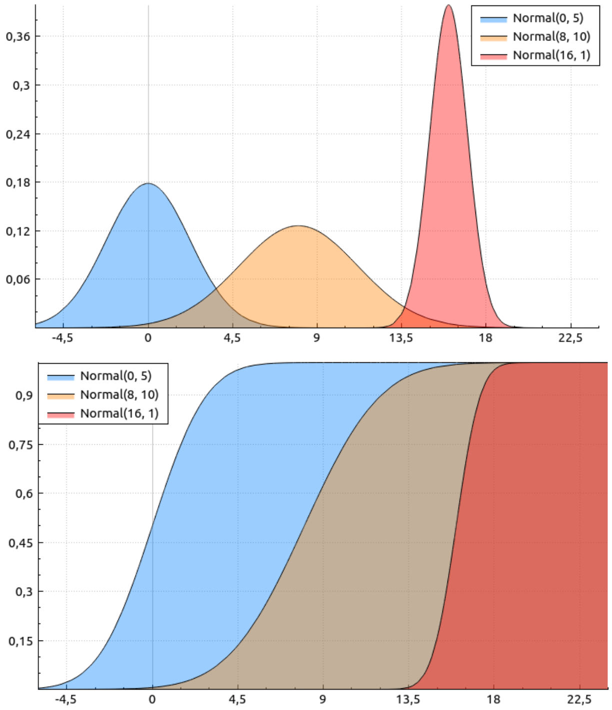
Relation to Stable distribution:

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

10.4 Holtsmark distribution

Relation to Stable distribution:

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



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

10.5 Landau distribution

Relation to Stable distribution:

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

11 Pareto distribution

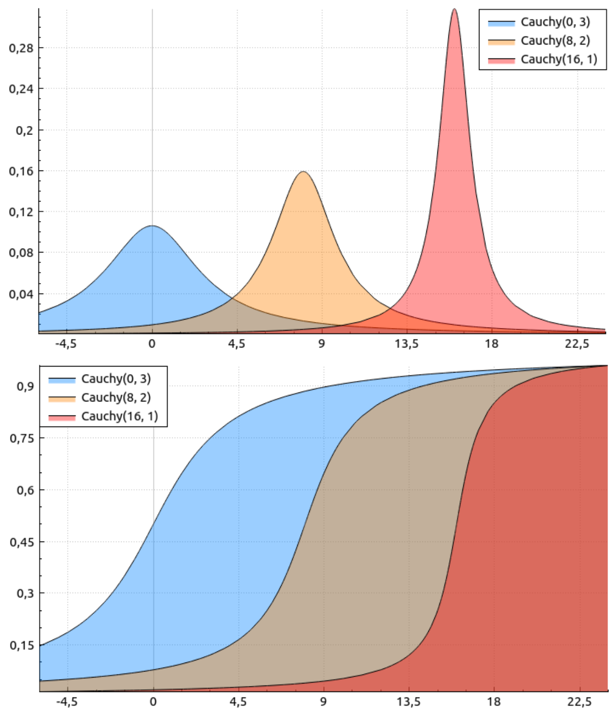
Estimation of parameters.

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:

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



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

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

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

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

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

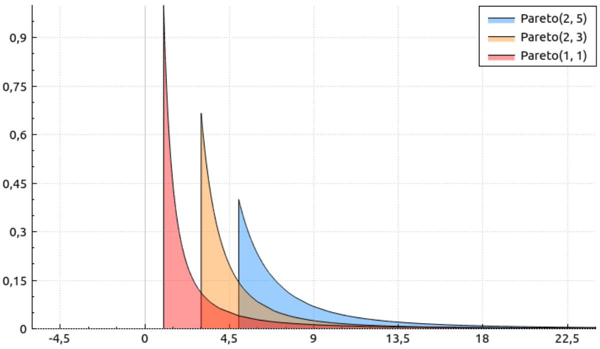
Therefore, in order to get unbiased estimators we need to make the following transformations:

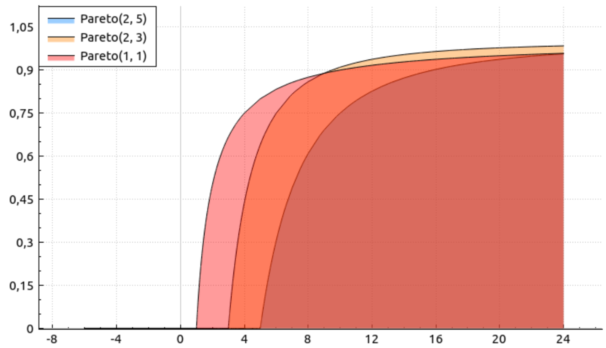
$$\tilde{\alpha} = \frac{n-2}{n}\hat{\alpha} \quad \text{and} \quad \tilde{\sigma} = \hat{\sigma}\left(1 - \frac{1}{(n-1)\hat{\alpha}}\right).$$

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 $\Gamma(\kappa, \beta)$:

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





Notation	$X \sim \text{Pareto}(\alpha, \sigma)$
Parameters	$\alpha, \sigma > 0$
Support	$x \geq \sigma$
$f(x)$	$\frac{\alpha\sigma^\alpha}{x^{\alpha+1}}$
$F(x)$	$1 - \left(\frac{\sigma}{x}\right)^\alpha$
$\mathbb{E}[X]$	$\frac{\alpha\sigma}{\alpha-1}$ 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

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:

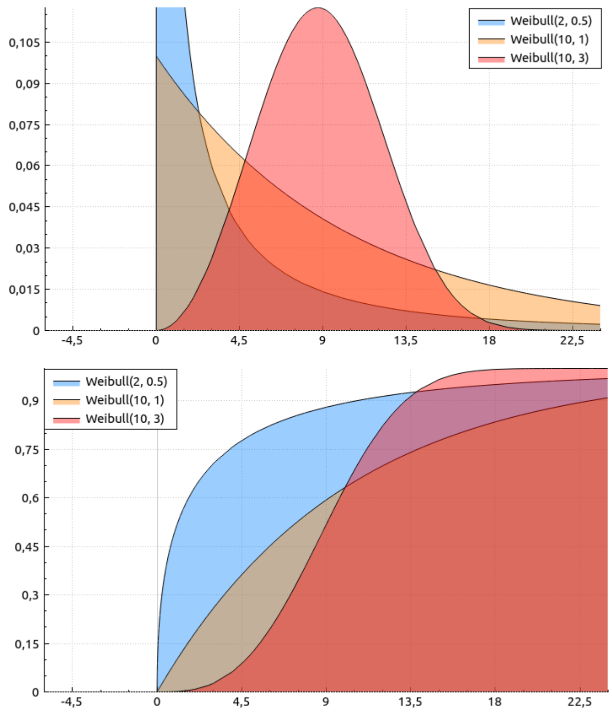
$$\mathbb{E}[\sigma|X] = \frac{\theta^{\alpha'}}{1 - \left(\frac{\theta}{X_{(1)}}\right)^{\alpha'}} \cdot \left(\frac{\alpha'}{\alpha' - 1}\right) \cdot \left(\frac{1}{\theta^{\alpha'}} - \frac{1}{X_{(1)}^{\alpha'}}\right)$$

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

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

However, Bounded-Pareto distribution is not yet supported.

12 Weibull



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

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

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 Inv-$\Gamma(\alpha, \beta)$:

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

Posterior distribution then:

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

Bayesian estimator:

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

MAP estimator:

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

Part III
Discrete univariate distributions

13 Beta-binomial distribution

14 Binomial distribution

Notation	$X \sim \mathrm{Bin}(n, p)$
Parameters	$n \in \mathbb{N}, p \in [0, 1]$
Support	$k \in \{0, \ldots, n\}$
P.m.f.	$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
$F(x)$	$\mathbb{P}(X \leq k) = I_{1-p}(n-k, 1+k)$
$\mathbb{E}[X]$	np
$\mathrm{Var}(X)$	$np(1-p)$
Median	$[np]$
Mode	$[(n+1)p]$
$\phi(t)$	$(1 - p + pe^{it})^n$

14.1 Bernoulli

Notation:

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

Relation to Binomial distribution:

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

15 Poisson distribution

Notation	$X \sim \mathrm{Po}(\lambda)$
Parameters	$\lambda > 0$
Support	$k \in \mathbb{N}_0$
P.m.f.	$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
$F(x)$	$\mathbb{P}(X \leq k) = Q(k+1, \lambda)$
$\mathbb{E}[X]$	λ
$\mathrm{Var}(X)$	λ
Median	$\sim \max\big([\lambda + \frac{1}{3} - \frac{0.02}{\lambda}], 0\big)$
Mode	$[\lambda]$
$\phi(t)$	$\exp\{\lambda(e^{it} - 1)\}$

Generator (let $\delta = \mu \in \mathbb{Z}$). (There is a mistake in Lemma 3.8 in first inequality). Recall that

$$q(X) = X \ln(\lambda) - \ln\Big(\frac{(\mu + X)!}{\mu!}\Big).$$

We denote acceptance probability $\mathbb{P}(W \leq q(X))$ by p.

- $k = \mu$. Probability to be in this setting is $1/c$.

$$\mathbb{P}(X = 0|W \leq q(X)) = \frac{\mathbb{P}(X = 0, W \leq q(X))}{\mathbb{P}(W \leq q(X))} = \frac{1}{pc}.$$

  On the other hand it should be equal to:

$$\frac{1}{pc} = \frac{\lambda^{\mu} e^{-\lambda}}{\mu!}.$$

- $k = \mu + 1$.

$$\mathbb{P}(X = 1|W \leq q(X)) = \frac{\mathbb{P}(X = 1, W \leq q(X))}{\mathbb{P}(W \leq q(X))} = \frac{\lambda}{p(\mu + 1)c}$$
$$= \frac{\lambda^{\mu+1} e^{-\lambda}}{(\mu + 1)!}.$$

- $k < \mu$. Here was mistake in the book. We adjust the probabilities. Probability to be in this setting is $\sqrt{\pi\mu/2e}/c$.

$$
\begin{aligned}
\mathbb{P}(W \leq q(X), X = k-\mu | U \leq c_1) &= \mathbb{P}\Bigg(-\frac{N^2}{2} + \frac{1}{2} - E < q(\lfloor -|N|\sqrt{\mu} \rfloor), \lceil |N|\sqrt{\mu} \rceil = \mu - k \Bigg) \\
&= \mathbb{P}\Bigg(-\frac{N^2}{2} + \frac{1}{2} - E < \lfloor -|N|\sqrt{\mu} \rfloor \ln(\lambda) - \ln\left(\frac{(\mu + \lfloor -|N|\sqrt{\mu} \rfloor)!}{\mu!} \right), \\
&\qquad \frac{\mu-k-1}{\sqrt{\mu}} \leq |N| < \frac{\mu-k}{\sqrt{\mu}} \Bigg) \\
&= \mathbb{P}\Bigg(U < \exp\left\{ \frac{N^2}{2} - \frac{1}{2} + \lfloor -|N|\sqrt{\mu} \rfloor \ln(\lambda) - \ln\left(\frac{(\mu + \lfloor -|N|\sqrt{\mu} \rfloor)!}{\mu!} \right.\right. \\
&\qquad \frac{\mu-k-1}{\sqrt{\mu}} \leq |N| < \frac{\mu-k}{\sqrt{\mu}} \Bigg) \\
&= \sqrt{\frac{2}{e\pi}} \int_{\frac{\mu-k-1}{\sqrt{\mu}}}^{\frac{\mu-k}{\sqrt{\mu}}} \exp\left\{ \lfloor -|n|\sqrt{\mu} \rfloor \ln(\lambda) - \ln\left(\frac{(\mu + \lfloor -|n|\sqrt{\mu} \rfloor)!}{\mu!} \right) \right\} dn \\
&= \sqrt{\frac{2}{e\pi\mu}} \int_{\mu-k-1}^{\mu-k} \exp\left\{ \lfloor -z \rfloor \ln(\lambda) - \ln\left(\frac{(\mu + \lfloor -z \rfloor)!}{\mu!} \right) \right\} dz \\
&= \sqrt{\frac{2}{e\pi\mu}} \exp\left\{ (k-\mu)\ln(\lambda) - \ln\left(\frac{k!}{\mu!} \right) \right\} \\
&= \sqrt{\frac{2}{e\pi\mu}} \lambda^{k-\mu} \frac{\mu!}{k!}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbb{P}(X = k-\mu | W \leq q(X)) &= \frac{\mathbb{P}(W \leq q(X), X = k-\mu)}{\mathbb{P}(W \leq q(X))} \\
&= \sqrt{\frac{2}{\pi\mu e}} \lambda^{k-\mu} \frac{\mu!}{k!} \cdot \sqrt{\pi\mu e/2} \frac{\lambda^{\mu} e^{-\lambda}}{\mu!} \\
&= \frac{\lambda^k e^{-\lambda}}{k!}
\end{aligned}
$$

- $k \in [\mu+2, 2\mu]$. Probability to be in this setting is $\sqrt{\frac{3\pi\mu}{4}} e^{\frac{1}{3\mu}}/c$. We also have

$$
W = \frac{-Y^2 + 2Y}{3\mu} - E = \frac{1}{3\mu} - \frac{N^2}{2} - E.
$$

Then

$$
\begin{aligned}
\mathbb{P}(W \le q(X)|X = k-\mu|U \in ...) &= \mathbb{P}\left(\frac{1}{3\mu} - \frac{N^2}{2} - E < q(\lceil 1 + |N|\sqrt{3\mu/2}\rceil), \lceil 1 + |N|\sqrt{3\mu/2}\rceil = k - \mu\right) \\
&= \mathbb{P}\Bigg(U < \exp\left\{-\frac{1}{3\mu} + \frac{N^2}{2} + q(\lceil 1 + |N|\sqrt{3\mu/2}\rceil)\right\}, \\
&\qquad \frac{k-\mu-2}{\sqrt{3\mu/2}} < |N| \le \frac{k-\mu-1}{\sqrt{3\mu/2}}\Bigg) \\
&= \sqrt{\frac{2}{\pi}} e^{-\frac{1}{3\mu}} \int_{\frac{k-\mu-2}{\sqrt{3\mu/2}}}^{\frac{k-\mu-1}{\sqrt{3\mu/2}}} \exp\left\{q(\lceil 1 + |n|\sqrt{3\mu/2}\rceil)\right\} dn \\
&= \sqrt{\frac{4}{3\pi\mu}} e^{-\frac{1}{3\mu}} \int_{k-\mu-1}^{k-\mu} \exp\left\{\lceil z \rceil \ln(\lambda) - \ln\left(\frac{(\mu + \lceil z \rceil)!}{\mu!}\right)\right\} dz \\
&= \sqrt{\frac{4}{3\pi\mu}} e^{-\frac{1}{3\mu}} \mu! \frac{\lambda^{k-\mu}}{k!}.
\end{aligned}
$$

- $k > 2\mu$. Probability to be in this setting is $6e^{-\frac{2+\mu}{6}}/c$.

$$
\begin{aligned}
\mathbb{P}(W \le q(X)|X = k-\mu|U \in ...) &= \mathbb{P}\left(-\frac{2+\mu}{6} - V - E < q(\lceil \mu + 6V \rceil), \lceil \mu + 6V \rceil = k - \mu\right) \\
&= \mathbb{P}\Bigg(-\frac{2+\mu}{6} - V + \ln(U) < \lceil \mu + 6V \rceil \ln(\lambda) - \ln\left(\frac{(\mu + \lceil \lambda + 6V \rceil)!}{\mu!}\right) \\
&\qquad \lceil \mu + 6V \rceil = k - \mu\Bigg) \\
&= \mathbb{P}\Bigg(U < \exp\left\{\frac{2+\mu}{6} + V + \lceil \mu + 6V \rceil \ln(\lambda) - \ln\left(\frac{(\mu + \lceil \mu + 6V \rceil)!}{\mu!}\right)\right\} \\
&\qquad \frac{k-2\mu-1}{6} < V \le \frac{k-2\mu}{6}\Bigg) \\
&= \int_{\frac{k-2\mu-1}{6}}^{\frac{k-2\mu}{6}} \exp\left\{\frac{2+\mu}{6} + \lceil \mu + 6v \rceil \ln(\lambda) - \ln\left(\frac{(\mu + \lceil \mu + 6v \rceil)!}{\mu!}\right)\right\} dv \\
&= \frac{e^{\frac{2+\lambda}{6}}}{6} \int_{k-\mu-1}^{k-\mu} \exp\left\{\lceil z \rceil \ln(\lambda) - \ln\left(\frac{(\mu + \lceil z \rceil)!}{\mu!}\right)\right\} dz \\
&= \frac{e^{\frac{2+\lambda}{6}}}{6} \exp\left\{(k-\mu)\ln(\lambda) - \ln\left(\frac{k!}{\mu!}\right)\right\} \\
&= \frac{e^{\frac{2+\lambda}{6}}}{6} \lambda^{k-\mu} \frac{\mu!}{k!}
\end{aligned}
$$

Hence,

$$\begin{aligned}
\mathbb{P}(X = k - \mu | W \leq q(X)) &= \frac{\mathbb{P}(W \leq q(X), X = k - \mu)}{\mathbb{P}(W \leq q(X))} \\
&= \frac{e^{\frac{2+\lambda}{6}}}{6} \lambda^{k-\mu} \frac{\mu!}{k!} \cdot \frac{6e^{-\frac{2+\mu}{6}}}{pc} \\
&= \frac{\lambda^k e^{-\lambda}}{k!}
\end{aligned}$$

Part IV

Bivariate distributions

16 Bivariate Normal distribution

17 Normal-Inverse-Gamma distribution

18 Trinomial distribution

Part V
Circular distributions

19 von Mises distribution

20 Wrapped Exponential distribution

Part VI

Singular distributions

21 Cantor distribution