

Quantile-parameterized distributions for expert knowledge elicitation

Supplementary Materials

Dmytro Perepolkin

Erik Lindström

Ullrika Sahlin

Appendix A. Distribution functions

Myerson Distribution

The derivative of the quantile function with respect to the depth u is the Quantile Density Function, which for Myerson distribution has the following form

$$q(u|q_1, q_2, q_3, \alpha) = \begin{cases} \rho \frac{\beta^\kappa \ln(\beta)}{(\beta-1)} \frac{q_{norm}(u)}{\Phi^{-1}(1-\alpha)}, & \beta \neq 1 \\ \rho \frac{q_{norm}(u)}{\Phi^{-1}(1-\alpha)}, & \beta = 1 \end{cases}$$

where $q_{norm} = \frac{d\Phi^{-1}(u)}{du}$ is the quantile density function for the standard normal distribution.

The Myerson distribution is invertible. The distribution function of random variable X has the form

$$\begin{aligned} \psi &= \Phi^{-1}(1-\alpha) \left(\frac{\ln \left(1 + \frac{(x-q_2)(\beta-1)}{\rho} \right)}{\ln(\beta)} \right) \\ F(x|q_1, q_2, q_3, \alpha) &= \\ &\begin{cases} \Phi(\psi), & \beta \neq 1 \\ F_{norm}(x|q_2, \rho/\Phi^{-1}(1-\alpha)), & \beta = 1 \end{cases} \end{aligned}$$

where $\Phi()$ is the CDF of the standard normal distribution and $\Phi^{-1}()$ is its inverse. $F_{norm}(x|q_2, \rho/\Phi^{-1}(1-\alpha))$ is the CDF of the normal distribution with mean $\mu = q_2$ and standard deviation $\sigma = \rho/\Phi^{-1}(1-\alpha)$.

The derivative of the distribution function with respect to the random variable X is the probability density function, which for the Myerson distribution takes the following form

$$f(x|q_1, q_2, q_3, \alpha) = \begin{cases} \frac{\Phi^{-1}(1-\alpha)(\beta-1)}{(\rho+(x-q_2)(\beta-1))\ln(\beta)} \varphi(\psi), & \beta \neq 1 \\ f_{normal}(x|q_2, \rho/\Phi^{-1}(1-\alpha)), & \beta = 1 \end{cases}$$

where $\varphi()$ is the probability density function of the standard normal distribution, $f_{normal}(x|q_2, \frac{\rho}{\Phi^{-1}(1-\alpha)})$ is the PDF of the normal distribution with the mean $\mu = q_2$ and standard deviation $\sigma = \rho/\Phi^{-1}(1-\alpha)$.

Generalized Myerson Distributions

The Quantile Density Function of Generalized Myerson Distribution for $u \neq 0, u \neq 1$ is

$$q_M(u|q_1, q_2, q_3, \alpha) = \begin{cases} \rho \frac{\beta^\kappa \ln(\beta)}{(\beta-1)} \frac{s(u)}{S(1-\alpha)}, & \beta \neq 1 \\ \rho \frac{s(u)}{S(1-\alpha)}, & \beta = 1 \end{cases}$$

where $S(u)$ is the quantile function and $s(u) = \frac{dS(u)}{du}$ is the quantile density function for the kernel distribution. When $u = 0$ or $u = 1$ the $q_M(u) = \infty$.

The Generalized Myerson distribution is invertible. The distribution function of random variable X has the form

$$\psi = S(1-\alpha) \left(\frac{\ln \left(1 + \frac{(x-q_2)(\beta-1)}{\rho} \right)}{\ln(\beta)} \right)$$

$$F_M(x|q_1, q_2, q_3, \alpha) = \begin{cases} F(\psi), & \beta \neq 1 \\ q_2 + \frac{\rho}{S(1-\alpha)} F(x), & \beta = 1 \end{cases}$$

where $F()$ is the standard CDF of the kernel distribution and $S()$ is its inverse.

The derivative of the distribution function with respect to the random variable X is the probability density function, which for the Myerson distribution takes the following form

$$f_M(x|q_1, q_2, q_3, \alpha) = \begin{cases} \frac{S(1-\alpha)(\beta-1)}{(\rho+(x-q_2)(\beta-1))\ln(\beta)} f(\psi), & \beta \neq 1 \\ f\left(\frac{x-q_2}{\rho/S(1-\alpha)}\right), & \beta = 1 \end{cases}$$

where $f()$ is the probability density function of the standard kernel distribution. Compare it to the simplicity of the Quantile Density Function above.

Johnson Quantile-Parameterized Distribution

The JQPD-B quantile density function can be computed as

$$q_B(p) = \begin{cases} (u_b - l_b)\varphi(\xi + \lambda \sinh(\delta(z(p) + nc))) \times \\ \quad \times \lambda \cosh(\sigma(z(p) + nc))\sigma q_{norm}(p), & n \neq 0 \\ (u_b - l_b)\varphi\left(B + \left(\frac{H-L}{2c}\right)z(p)\right) \times \\ \quad \times \left(\frac{H-L}{2c}\right)q_{norm}(p), & n = 0 \end{cases}$$

The JQPD-B distribution function

$$F_B(x) = \begin{cases} \Phi\left((2c/(H-L))(-B + z\left(\frac{x-l}{u-l}\right))\right), & n = 0 \\ \Phi\left(\frac{1}{\delta} \sinh^{-1}\left(\frac{1}{\lambda}\left(z\left(\frac{x-l}{u-l}\right) - \xi\right) - nc\right)\right), & n \neq 0 \end{cases}$$

The JQPD-B probability density function (PDF) is

$$f(x) = \begin{cases} \frac{2c}{(H-L)(u_b-l_b)} \frac{1}{\varphi\left(z\left(\frac{x-l_b}{u_b-l_b}\right)\right)} \varphi\left(\frac{2c}{H-L}\left(-B + z\left(\frac{x-l_b}{u_b-l_b}\right)\right)\right), & n = 0 \\ \frac{1}{\delta} \frac{1}{u_b-l_b} \varphi\left(-nc + \frac{1}{\delta} \sinh^{-1}\left(\frac{1}{\lambda}\left(-\xi + z\left(\frac{x-l_b}{u_b-l_b}\right)\right)\right)\right) \times \\ \quad \frac{1}{\varphi\left(z\left(\frac{x-l_b}{u_b-l_b}\right)\right)} \frac{1}{\sqrt{\lambda^2 + \left(-\xi + z\left(\frac{x-l_b}{u_b-l_b}\right)\right)^2}}, & n \neq 0 \end{cases}$$

J-QPD-S quantile density function

$$q_S(p) = \begin{cases} \theta \exp(\lambda \delta z(p)) \lambda \delta q_{norm}(p), & n = 0 \\ \theta \exp\left(\lambda \sinh^{-1}(\delta z(p)) + \sinh^{-1}(nc\delta)\right) \lambda \frac{1}{\sqrt{1+(\delta z(p))^2}} \delta q_{norm}(p), & n \neq 0 \end{cases}$$

J-QPD-S distribution function

$$F_S(x) = \begin{cases} F_{lnorm}(x - l_b | \ln(\theta), \frac{H-B}{c}), & n = 0 \\ \Phi\left(\frac{1}{\delta} \sinh\left(\sinh^{-1}\left(\frac{1}{\lambda} \ln \frac{x-l_b}{\theta}\right) - \sinh^{-1}(nc\delta)\right)\right), & n \neq 0 \end{cases}$$

J-QPD-S probability density function (PDF)

$$f_S(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \ln \xi)^2}{2\frac{(H-B)^2}{c^2}}\right), & n = 0 \\ \varphi\left(\frac{\sinh(\sinh^{-1}(cn\sigma) - \sinh^{-1}(\frac{1}{\lambda} \ln \frac{x-l_b}{\theta}))}{\delta}\right) \frac{\cosh(\sinh^{-1}(cn\delta) - \sinh^{-1}(\frac{1}{\lambda} \ln \frac{x-l_b}{\theta}))}{(x-l_b)\delta\lambda\sqrt{1+\left(\frac{\ln \frac{x-l_b}{\theta}}{\lambda}\right)^2}}, & n \neq 0 \end{cases}$$

where $\mu = \ln \xi$ and $\sigma = \frac{H-B}{c}$.

Metalog distribution

This section recapitulates ideas and formulas provided in (Keelin 2016) with our own notation and minor reinterpretations.

Metalog distribution is created from the logistic quantile function $Q(p) = \mu + s \text{logit}(p)$, where μ is the mean, s is proportional to the standard deviation such that $\sigma = s\pi/\sqrt{3}$, p is the probability $p \in [0, 1]$. The metalog quantile function is built by substitution and series expansion of its parameters μ and s with the polynomial of the form:

$$\begin{aligned} \mu &= a_1 + a_4(p - 0.5) + a_5(p - 0.5)^2 + a_7(p - 0.5)^3 + a_9(p - 0.5)^4 + \dots, \\ s &= a_2 + a_3(p - 0.5) + a_6(p - 0.5)^2 + a_8(p - 0.5)^3 + a_{10}(p - 0.5)^4 + \dots, \end{aligned}$$

where a_i , $i \in (1 \dots n)$ are real constants. Given a size- m QPT $\{p, q\}_m$, where $p = \{p_1 \dots p_m\}$ and $q = \{q_1 \dots q_m\}$ the vector of coefficients $a = \{a_1 \dots a_m\}$ can be determined through the set of linear equations.

$$\begin{aligned} q_1 &= a_1 + a_2 \text{logit}(p_1) + a_3(p_1 - 0.5) \text{logit}(p_1) + a_4(p_1 - 0.5) + \dots, \\ q_2 &= a_1 + a_2 \text{logit}(p_2) + a_3(p_2 - 0.5) \text{logit}(p_2) + a_4(p_2 - 0.5) + \dots, \\ &\vdots \\ q_m &= a_1 + a_2 \text{logit}(p_m) + a_3(p_m - 0.5) \text{logit}(p_m) + a_4(p_m - 0.5) + \dots. \end{aligned}$$

In the matrix form, this system of equations is equivalent to $q = \mathbb{P}a$, where q and a are column vectors and \mathbb{P} is a $m \times n$ matrix:

$$\mathbb{P} = \begin{bmatrix} 1 & \text{logit}(p_1) & (p_1 - 0.5) \text{logit}(p_1) & (p_1 - 0.5) & \dots \\ 1 & \text{logit}(p_2) & (p_2 - 0.5) \text{logit}(p_2) & (p_2 - 0.5) & \dots \\ & & \vdots & & \\ 1 & \text{logit}(p_m) & (p_m - 0.5) \text{logit}(p_m) & (p_m - 0.5) & \dots \end{bmatrix}$$

If $m = n$ and \mathbb{P} is invertible, then the vector of coefficients a of this *properly parameterized* metalog QPD can be uniquely determined by

$$a = \mathbb{P}^{-1}q \quad (1)$$

If $m > n$ and \mathbb{P} has a rank of at least n , then the vector of coefficients a of the *approximated* metalog QPD, can be estimated using

$$a = [\mathbb{P}^T \mathbb{P}]^{-1} \mathbb{P}^T q$$

The matrix to be inverted is always $n \times n$ regardless of the size m of QPT used.

Metalog *quantile function* (QF) with n terms $Q_{M_n}(u|a)$ can be expressed as

$$Q_{M_n}(u|a) = \begin{cases} a_1 + a_2 \text{logit}(u), & \text{for } n = 2, \\ a_1 + a_2 \text{logit}(u) + a_3(u - 0.5) \text{logit}(u), & \text{for } n = 3, \\ a_1 + a_2 \text{logit}(u) + a_3(u - 0.5) \text{logit}(u) + a_4(u - 0.5), & \text{for } n = 4, \\ Q_{M_{n-1}} + a_n(u - 0.5)^{(n-1)/2}, & \text{for odd } n \geq 5, \\ Q_{M_{n-1}} + a_n(u - 0.5)^{n/2-1} \text{logit}(u), & \text{for even } n \geq 6, \end{cases} \quad (2)$$

where $u \in [0, 1]$ is the cumulative probability and a is the size- n parameter vector of real constants $a = \{a_1 \dots a_n\}$.

The metalog *quantile density function* (QDF) can be found by differentiating the Equation 2 with respect to u :

$$q_{M_n}(u|a) = \begin{cases} a_2 \mathcal{J}(u), & \text{for } n = 2, \\ a_2 \mathcal{J}(u) + a_3((u - 0.5) \mathcal{J}(u) + \text{logit}(u)), & \text{for } n = 3, \\ a_2 \mathcal{J}(u) + a_3((u - 0.5) \mathcal{J}(u) + \text{logit}(u)) + a_4, & \text{for } n = 4, \\ q_{M_{n-1}} + 0.5a_n(n-1)(u - 0.5)^{(n-3)/2}, & \text{for odd } n \geq 5, \\ q_{M_{n-1}} + a_n((u - 0.5)^{n/2-1} \mathcal{J}(u) + (0.5n - 1)(u - 0.5)^{n/2-2} \text{logit}(u)), & \text{for even } n \geq 6, \end{cases} \quad (3)$$

where $\mathcal{J}(u) = [u(1 - u)]^{-1}$. The constants a are feasible iff $q_{M_n}(u|a) > 0$, $\forall u \in [0, 1]$.

Metalog *density quantile function* (DQF), referred to as the “metalog pdf” in (Keelin 2016) can be obtained by $f(Q_{M_n}(u|a)) = [q_{M_n}(u|a)]^{-1}$.

Metalog *cumulative distribution function* (CDF) $F_{M_n}(x|a)$ does not have an explicit form because $Q_{M_n}(u|a)$ is not invertible (Keelin 2016). It is, however, possible to approximate $\widehat{Q}_{M_n}^{-1}(x|a)$ using approximation.

Metalog distribution is defined for all $x \in \mathbb{R}$ on the real line. (Keelin 2016) provides semi-bounded *log-metalog*, and the bounded *logit-metalog* variations of the metalog distribution. As

the names suggest, this is achieved through the variable substitution with $z = \ln(x - b_l)$ or $z = -\ln(b_u - x)$ for the semi-bounded case, and $z = \ln((x - b_l)/(b_u - x))$ for the bounded case, where z is metalog-distributed and b_l, b_u are the lower and upper limits, respectively. Substituting one of the transformations into the QF and QDF functions above, yields semi-bounded or bounded metalog distribution. For the exact formulae of the log-metalog and logit-metalog refer to (Keelin 2016).

CSW GLD

Quantile density function for the CSW GLD is provided in (Chalabi, Scott, and Wuertz 2012)

$$q(u|\tilde{\sigma}, \chi, \xi) = \frac{\tilde{\sigma}}{S(0.75|\chi, \xi) - S(0.25|\chi, \xi)} s(u|\chi, \xi)$$

$$s(u|\chi, \xi) = \frac{d}{du} S(u|\chi, \xi) = u^{\alpha+\beta-1} + (1-u)^{\alpha-\beta-1}$$

References

- Chalabi, Yohan, David J Scott, and Diethelm Wuertz. 2012. “Flexible Distribution Modeling with the Generalized Lambda Distribution.” Working Paper MPRA Paper No. 43333,. Zurich, Switzerland: ETH.
- Keelin, Thomas W. 2016. “The Metalog Distributions.” *Decision Analysis* 13 (4): 243–77. <https://doi.org/10.1287/deca.2016.0338>.