ECMOR XI

Asymptotic Dykstra-Parsons estimates and confidence intervals

Salvador Pintos, Claritza Bohorquez, and Nestor V. Queipo

Applied Computing Institute, Faculty of Engineering, University of Zulia, Venezuela

{spintos, claritza, nqueipo}@ica.luz.ve

Abstract

The Dykstra-Parson (DP), the most popular heterogeneity static measure among petroleum engineers, may be at a significant error, in particular when assumptions are made about the permeability distribution (parametric approaches) that may lead to unrealistic reservoir performance predictions and unsuccessful development plans. This paper presents the development of an asymptotic distribution of the Dykstra-Parsons coefficient that is independent of the probability distribution of the permeability variable. The effectiveness (bias and confidence intervals) of the proposed approach is demonstrated by comparing the results those obtained using the classical method, and well-known parametric methods, under different scenarios of reservoir maturity levels (i.e., number of wells), and degree violations of the log-Normal probability density function assumption. The results show that in the vast majority of the case studies the proposed approach outperformed previously reported methods, in particular, resulted in a significant reduction of the bias and, with confidence intervals always including the estimated DP coefficient. In addition, an excellent agreement was observed between the asymptotic cumulative distribution of the DP coefficient and the corresponding empirical distribution for sample sizes as low as 100, which allows classifying reservoirs according to their DP coefficient with high success rates.

I. Introduction

In the context of enhanced oil recovery projects, heterogeneity (the spatial variation of properties) has long been recognized as a key component in predicting reservoir performance, namely, amount of petroleum recovered, time to breakthrough, and peak hydrocarbon production (Jensen et al., 1986; Jensen and Lake, 1986; Lake and Jensen, 1989; Jensen and Currie, 1990). While the complexity of the heterogeneity/performance relationship is well documented, for screening purposes, or to establish if a more detailed study is justified, the Dykstra-Parsons (DP) coefficient remains as the most popular heterogeneity static measure among petroleum engineers. The DP coefficient estimates, though, may be at a significant error, so assessing its bias and confidence intervals are issues of considerable interest that may lead to more realistic reservoir performance predictions and successful development plans.

Previous works based for DP estimates and confidence intervals assume the reservoir property of interest (typically, permeability) has associated a log-Normal probability density function or there is a transformation (Box-Cox) that can lead to Normal behavior (Jensen and Lake, 1986); frequently, however, this assumption does not hold. *This paper presents a novel approach that overcome this limitation and provides sample DP estimates and confidence intervals based on the asymptotic Normal behavior of the joint probability distribution of cuantiles.*

The following section discusses the theoretical and sampled DP coefficient, Sections III presents a frequently used parametric approach, Section IV develops an asymptotic DP estimator, a description of the case studies used for evaluating the relative performance of the proposed approach is the subject of Section V, while Section VI and VII discuss the results obtained, and the most significant conclusions, respectively.

II. Dykstra-Parsons coefficient

It represents a *robust estimation* of the well-known coefficient of variation σ/μ, a normalized measure of dispersion of a probability distribution, used for describing reservoir permeability heterogeneity. Dykstra y Parsons (1950) recognized that the classical coefficient of variation for asymmetric probability distributions (such as those associated to permeability) was very sensitive to extreme values, in particular, considering the relatively small sample sizes typically available in oil industry environments. As a result, they substituted the classical statistics σ and μ for analog quantities calculated using order statistics (quantiles) of the probability distribution of reservoir permeability. More specifically, if the probabilities $p_1 = 0.1587$, and $p_2 = 0.5$, and F is the cumulative probability distribution of the permeability of interest (random variable X), the theoretical Dykstra-Parsons (DP_T) coefficient is calculated as:

$$DP_T = \frac{F^{-1}(p_2) - F^{-1}(p_1)}{F^{-1}(p_2)} = \frac{xp_2 - xp_1}{xp_2}$$

where xp_i is the quantile of the probability distribution of X associated with p_i, that is, $F(xp_i)=p_i$. The probabilities $p_1 = 0.1587$, and $p_2 = 0.5$ are such that if the cited probability distribution is normal the DP_T is equivalent to the coefficient of variation. Note that xp_2 is the median of the population and that for positive random variables such as permeability, $0<DP_T<1$.

Since only a sample (size n) of the random variable X is available, $X_1,......,X_n$, this is ordered such as, $X_{(1)} <= X_{(2)} ....<= X_{(n)}$, and each element $X_{(i)}$ represents the i-th order statistics. Then, the *sampled DP coefficient* (Equation 1) is calculated using sample quantiles $q_n(p)$ where, for a given probability p, $q_n(p)$ is the h-th order statistics $X_{(h)}$ with h = [np] + 1 corresponding to the sample size n; the symbol [.] denotes the integer part operator.

$$DP = \frac{q_n(0.5) - q_n(0.159)}{q_n(0.5)} = 1 - \frac{q_n(0.159)}{q_n(0.5)} \qquad (1)$$

III. Dykstra-Parsons coefficient estimator – Parametric scenario

Jensen and Currie (1990) show that if the probability distribution of the permeability (random variable X) is assumed to be Log-normal, it is possible to obtain a consistent DP coefficient estimator that outperforms (in terms of bias and variance) the corresponding sampled DP coefficient. Specifically, if the random variable X is $LogNormal(\mu,\sigma^2)$, the theoretical DP can be expressed as:

$$DP_T = 1 - \frac{e^{\mu-\sigma}}{e^{-\mu}} = 1 - e^{-\sigma}$$

which as expected grows with increasing values of σ. Since σ is unknown in the expression above, it is substituted by an unbiased estimator $w = \frac{s}{C4}$, where $C4 \cong 1 - \frac{1}{4n} - \frac{7}{32n^2}$, and *s* is the sampled (log of sample values) standard deviation.

The expected value of the Jensen and Currie (1990) estimator can then approximated as (using the Delta-Method) by:

$$E(DP_J) \cong 1 - e^{-\sigma} - e^{-\sigma}\frac{\sigma^2}{4n}$$

Note that the expression above underestimates the theoretical DP coefficient (DP_T) and the bias is proportional to the inverse of the sample size (n). The standard deviation of the Jensen and Currie estimator can be approximated as:

$$\sigma(DP_J) \cong e^{-\sigma}\frac{\sigma}{\sqrt{2n}}$$

Assuming the DP_J estimator is Normally distributed, Jensen and Currie set up a 95% confidence interval equal to:

$$DP_J \pm 2e^{-\sigma}\frac{\sigma}{\sqrt{2n}}$$

While the Log-normal assumption has been useful, it is well-known that, frequently, this assumption may not hold. See, for example, Lambert (1981), Goggin et al. (1986), and Jensen and Lake (1985).

IV. Dykstra-Parsons coefficient estimator – Non-parametric scenario

This section presents the development of an asymptotic distribution of the Dykstra-Parsons coefficient that is independent of the probability distribution of the permeability (random variable X).

Theorem (Cramer, 1999). The joint distribution of two centered quantiles denoted as $q_n(p_1)$ and $q_n(p_2)$ with ($p_1 < p_2$) is asymptotically Normal with mean a vector with the quantiles of the population:

$$\mu = \begin{bmatrix} xp_1 \\ xp_2 \end{bmatrix}$$

And convariance matrix COV/n, where:

$$COV = \begin{bmatrix} \dfrac{p_1(1-p_1)}{f^2(xp_1)} & \dfrac{p_1(1-p_2)}{f(xp_1)f(xp_2)} \\ \dfrac{p_1(1-p_2)}{f(xp_1)f(xp_2)} & \dfrac{p_2(1-p_2)}{f^2(xp_2)} \end{bmatrix}$$

The symbol f denotes the probability distribution associated with the random variable X. More precisely, when $n \to \infty$ the following expression converge to a Normal probability distribution,

$$n^{1/2}\begin{pmatrix} q_n(p_1) - xp_1 \\ q_n(p_2) - xp_2 \end{pmatrix} \Rightarrow N(0, COV)$$

Note the symmetric and positive definite nature of matrix COV and, the asymptotic distribution of the individual quantiles.

On the other hand, the probability distribution of a linear combination of quantiles (Z):

$$Z = aq_n(p_1) + bq_n(p_2)$$

can be derived. Following the previous result, Z is asymptotically Normal with expected value:

$$E(Z) = axp_1 + bxp_2 \qquad (2)$$

And variance $Var(Z) = [a\ b]COV(Z)\begin{bmatrix} a \\ b \end{bmatrix}$. More precisely:

$$Var(Z) = a^2 \frac{p_1(1-p_1)}{nf^2(xp_1)} + 2ab \frac{p_1(1-p_2)}{nf(xp_1)f(xp_2)} + b^2 \frac{p_2(1-p_2)}{nf^2(xp_2)} \qquad (3)$$

Hence, independently of the probability distribution of origin for the random variable X (permeability), the linear combination of quantiles Z follows a Normal probability distribution with mean and variance as specified by Equations 2 and 3.

Now, the cumulative probability distribution of the Dykstra-Parsons coefficient, $F_{DP}(x)$, can be estimated. The $F_{DPA}(x)$ is represented by:

$$F_{DPA}(y) = Prob(DP < y) = Prob\left(1 - \frac{q_n(0.159)}{q_n(0.5)} < y \right) \quad para\ \ 0 < y < 1$$

and can be rearranged as:

$$Prob(DP < y) = Prob(0 < q_n(0.159) + (y-1)q_n(0.5)) \qquad (4)$$

The expression in parenthesis in the right hand side of Equation 4, is a linear combination of quantiles as specified in Equation 2 with p_1=0.159, p_2=0.5, a=1 and b=y-1. Hence, Equation 4 can be rewritten as:

$$F_{DPA}(y) = Prob(DP < y) = Prob(0 < Z)$$

With variable Z Normally distributed with expected value:

$$E(Z) = xp_1 + (y-1)xp_2 \qquad (5)$$

where $xp_1 = F^{-1}(0.159)$ and $xp_2 = F^{-1}(0.5)$ are the quantiles required by the theoretical Dykstra-Parsons, DP_T. and, variance equal to:

$$Var(Z) = \frac{0.159*0.841}{nf^2(xp_1)} + 2(y-1)\frac{0.159*0.5}{nf(xp_1)f(xp_2)} + (y-1)^2 \frac{0.5*0.5}{nf^2(xp_2)} \qquad (6)$$

If the expected value E(Z) and square root of Var (Z) in Equations 5 and 6 are denoted by μ_z, and σ_z,, respectively, the asymptotic cumulative probability distribution of the Dykstra-Parsons coefficient can be written as:

$$F_{DPA}(y) = Prob(0 < Z) = 1 - Prob(Z < 0) = 1 - \Phi\left(\frac{-\mu_z}{\sigma_z}\right)$$

where $\Phi(.)$ denotes the cdf of the standard Normal distribution. Note that this expression depends only on the theoretical quantiles xp_1 and xp_2, the density function values $f(xp_1)$ y $f(xp_2)$ and sample size n. The cited theoretical quantiles and density function values could be approximated using the corresponding sample values.

Once the F_{DPA} distribution is available, the expected value of the asymptotic DP coefficient can be calculated as:

$$E(DP_A) = \int_0^1 (1 - F_{DPA}(y))dy$$

Since the $F_{DPA}(y)$ is available for every y, it is possible to numerically estimate the above-referenced integrals, as well as to establish the density function $f_{DPA}(y)$. Asymptotic confidence intervals [L,U] with a 1-α associated probability can also be obtained as $L = F_{DPa}^{-1}(\alpha/2)$ and $U = F_{DPA}^{-1}(1-\alpha/2)$. Note that the median of the asymptotic DP coefficient is equivalent to the DP_T since $F_{DPA}(y) = 0.5$ corresponds to $\mu_z=0$, and from $xp_1+(y-1)xp_2=0$ the DP coefficient is given by:

$$y = \frac{xp_2 - xp_1}{xp_2} = DP_T$$

V. Case studies: Two-way mixture of Log-normal probability distributions

The case studies were designed for frequent situations where an oil reservoir is represented by two litofacies and a mixture of Log-normal probability distributions may be a reasonable approximation to characterize the permeability. Different samples of this distribution under alternative scenarios of reservoir maturity levels (i.e., 100 and 200 wells) may or may not reject the Log-normal assumption (based on the p-value of the Lilliefors tests) which will help establish the relative performance of the parametric and asymptotic non-parametric approaches. In addition, the effectiveness of the proposed asymptotic DP estimator for decision making (e.g., classifying a reservoir as low, medium, or highly heterogeneous) is also evaluated.

The parameters of the two-way mixture of Log-normals are shown in Table 1 with the theoretical DP coefficient being equal to 0.667 (medium heterogeneity). The density function for this mixture constructed using Parzen windows are in general similar to single Log-normal probability distributions which may mislead partitioners on the true nature of the permeability statistical characterization.

Table 1. Parameters of the two-way mixture of Log-Normal probability distributions (Case studies)

Parameter	Distribution 1	Distribution 2
Permeability mean (µd)	600	145

Dykstra-Parsons coefficient	0.4	0.4
Weigh in the mixture (%)	60	40

VI. Results and discussion

With reference to Table 2, note that in general for samples of size 100, the DP_A estimates were close to the theoretical one, even in those instances where the Log-normal hypothesis could not be rejected. Furthermore, the median of the differences between the theoretical DP value and the DP-parametric and asymptotic estimates were approximately 0.08 (biased) and 0 (unbiased), respectively. Similar results were obtained for the larger sample size (200).

Table 2. Number of instances where the Dykstra-Parsons coefficients using the parametric and non-parametric approaches were closer to the theoretical DP coefficient (0.667).

Lilliefors Test – Null hipótesis (α=5%)	Sample size					
	100			200		
	Total	DP-parametric	DP-Asymptotic	Total	DP-parametric	DP-Asymptotic
Reject	58	4	54	93	2	91
Do not reject	42	1	41	7	0	7
Total	100	5	95	100	3	97

For a sample of size 100, Figures 1-a and 1-b exhibit the confidence intervals associated with the asymptotic DP and parametric estimates, respectively. The latter only included the theoretical value of the DP coefficient (0.667) in three instances, while the former included the theoretical DP coefficient in all but three occasions. Increasing the sample size to 200 led to DP parametric estimates farther away from the theoretical value.

The 95% confidence intervals for the asymptotic and parametric approaches were also compared using population parameters, for quantiles, and density function values, and, standard deviation, respectively. The lower and upper limits for the asymptotic DP confidence interval were (0.557 - 0.746), close to the theoretical one, namely, (0.573 - 0.747), while the confidence interval associated with the parametric DP approach (0.528 - 0.6311) is biased to the left and does not include the theoretical (0.667) Dykstra-Parsons coefficient. Similar results were obtained for the larger sample size.

Figure 2 shows the excellent agreement between the asymptotic cumulative probability distribution of the Dykstra-Parsons coefficient and the corresponding empirical cdf for 1000 simulations of sample sizes of 100 (a) and 400 (b). This agreement leads to the possibility of using the asymptotic $F_{DP}(x)$ to help support decisions related to whether to classify a reservoir as low, medium, or highly heterogeneous, for screening purposes, or to establish if a more detailed study is justified. As an example, for the case study where the DP coefficient is known (0.667), a common classifying scheme would make it a reservoir with medium heterogeneity (0.5<DP<0.7);

reservoirs with DP values higher than 0.7 would be considered as highly heterogeneous. Note the proximity of the known DP coefficient to the classifying frontier. A relevant question in this context would be what is the probability of an asymptotic DP coefficient higher than 0.7 (hence wrongly classifying it as highly heterogeneous) when in fact the DP coefficient is 0.667. Figure 3 shows the results corresponding to 10000 simulations and different sample sizes.

Note that even though the theoretical DP coefficient is close to the classification frontier (0.667 vs. 0.7) the average probability of an asymptotic Dykstra-Parsons coefficient higher than 0.7 is about 0.3 for a sample size of 100 and, as expected, it significantly decreases with larger sample sizes. Among the 10000 simulations, the actual samples with asymptotic Dykstra-Parsons coefficients higher than 0.7 were only about 20% (classification error) for the sample size equal to 100, and decreases to only about 9% and 4% for sample sizes of 300 and 500, respectively.
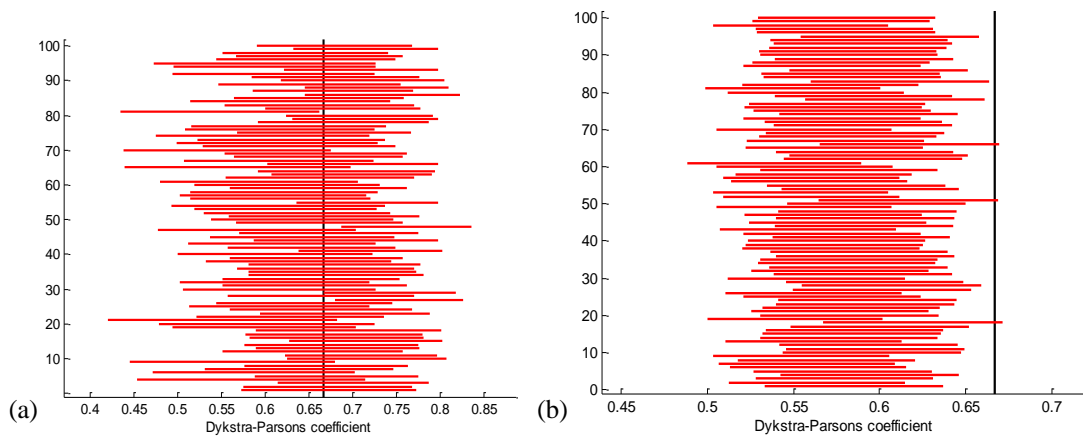

Figure 1. Confidence intervals associated with the asymptotic DP (a) and parametric (b) estimates for a sample of size 100. The vertical line represents the theoretical Dykstra-Parsons coefficient.
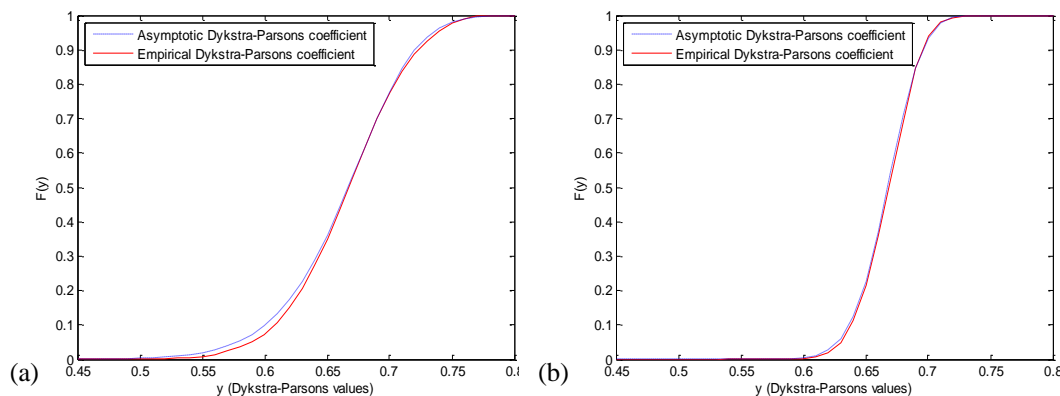

Figure 2. Asymptotic cumulative probability distribution of the Dykstra-Parsons coefficient and the corresponding empirical cdf for 1000 simulations of sample sizes of 100 (a) and 400 (b).

VII. Conclusions

This paper presents the development of an asymptotic distribution of the Dykstra-Parsons coefficient that is independent of the permeability probability distribution. The effectiveness (bias and confidence intervals) of the proposed approach is demonstrated by comparing the results those obtained using the classical method, and well-known methods (e.g., Jensen and Currie), under different scenarios of reservoir maturity levels (i.e., number of wells), and degree violations of the log-Normal probability density function assumption based on the p-value of Lilliefors tests.

The results show that in the vast majority of the case studies independently of whether or not the log-Normal probability density function assumption holds (with $\alpha=5\%$), the proposed approach outperformed previously reported methods, in particular, resulted in a significant reduction of the bias and, with confidence intervals always including the estimated DP coefficient. In addition, an excellent agreement was observed between the asymptotic cumulative distribution of the DP coefficient and the corresponding empirical distribution for sample sizes as low as 100, which allows classifying reservoirs according to their DP coefficient with high success rates.
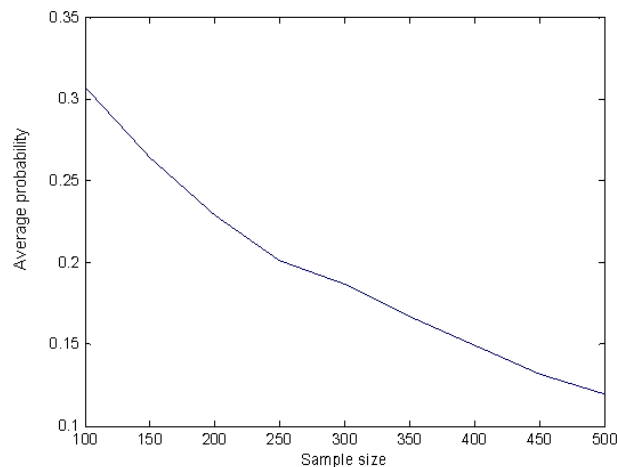

Figure 3. Average probability of an asymptotic Dykstra-Parsons coefficient higher than 0.7 (Case study)

The asymptotic distribution of the Dykstra-Parsons coefficient can be easily implemented as a computational aid and has the potential to be successfully incorporated in the workflow of reservoir engineers for quantifying/classifying reservoir heterogeneity without making any assumptions about the permeability probability distribution.

References

Cramer, H. [1999]. Mathematical methods of statistics. Princeton University Press. USA. 19th Edition.

Dykstra, H. and Parsons, R.L. [1950] The prediction of oil recovery by water flood, American Petroleum Institute, *Secondary Recovery of Oil in the United States,* second edition, API, Dallas, 160-174.

Goggin, D.J., Chandeler, M.A., Kocurek, G., and Lake, L.W. [1986] Patterns of permeability in eolian deposits, SPE 14893, *SPE/DOE 5th Symposium on EOR*, Tulsa, 297-306.

Jensen, J.L., Lake, L.W. and Hinkley, D.V. [1985] A statistical study of reservoir permeability: distributions, correlations and averages. SPE 14270, *The 60th SPE Annual Technical Conference and Exhibition*, Las Vegas, 461-468.

Jensen, J.L. and Lake, L.W. [1986] The influence of sample size and permeability distribution on heterogeneity measures. *The 61st SPE Annual Technical Conference and Exhibition*, Paper SPE 15434, Los Angeles.

Jensen J.L. and Currie, I. [1990] A new method for estimating the Dykstra-Parsons coefficient to characterize reservoir heterogeneity, SPE 17364, *SPE Reservoir Engineering*, **3**(7), 369-374.

Lake L.W. and Jensen, J.L. [1989] A review of heterogeneity measures used in reservoir characterization, SPE 20156-MS.

Lambert, M.E. [1981]: A statistical study of reservoir heterogeneity, MS Thesis, University of Texas at Austin.