Journal "Template"
. . . (20??)

Some Applications of Probability Generating Function based Methods to Statistical Estimation

Manuel L. Esquível

Departamento de Matemática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa Portugal
email: mle@fct.unl.pt

This work is dedicated to my dearest friend and colleague João Tiago Mexia, as a token of everlasting admiration, respect and gratitude, being certain that his unbreakable enthusiasm for Mathematics and mathematicians will always be, for us, a source of inspiration and guidance.

Abstract:

After recalling previous work on probability generating functions for real valued random variables we extend to these random variables uniform laws of large numbers and functional limit theorem for the empirical probability generating function. We present an application to the study of continuous laws, namely, estimation of parameters of Gaussian, gamma and uniform laws by means of a minimum contrast estimator that uses the empirical probability generating function of the sample. We test the procedure by simulation and we prove the consistency of the estimator.

Keywords: probability generating function, empirical laws, estimation of parameters of continuous laws

2000 Mathematics Subject Classification: 60–08, 60E10, 30B50

1 Introduction

Probability generating functions (PGF) computational techniques are often used for studying integer valued discrete random variables.

Recently, several works quoted in the references and specifically in section 4.3, developed interesting applications of these techniques for a wide range of purposes such as preliminary data analysis, estimation, tests, etc.

A detailed study of conditions for existence of probability generating functions for discrete random variables in [3] allows the extension of the available methods for integer or rational valued discrete random variables to real valued discrete random variables.

In this work we show that parameter estimation for continuous probability laws admitting non trivial PGF (such as Gaussian and gamma laws) may be performed using their PGF. Weaker versions of the results in this work were presented in [4].

2 PGF for real discrete random variables

In what follows and unless explicitly stated otherwise, X denotes a discrete random variable, $(\alpha_k)_{k\in\mathbb{Z}}$ being the real valued sequence of its values. With no generality loss we suppose that $\alpha_k < 0$ for $k < 0$, $\alpha_0 = 0$ e $\alpha_k > 0$ for $k > 0$.

For a sequence of non negative numbers $(p_k)_{k\in\mathbb{Z}}$ such that $\sum_{k=-\infty}^{+\infty} p_k = 1$, the probabilities, we have that $\mathbb{P}[X = \alpha_k] = p_k$. The PGF of X is

$$\psi_X(t) = \mathbb{E}[t^X] = \sum_{k=-\infty}^{+\infty} p_k t^{\alpha_k}$$

for $t > 0$. The natural domain of this PGF, $\mathbb{D}_X = \{t > 0 : \psi_X(t) < +\infty\}$ is clearly described in the following result (see [3]).

Theorem 2.1. *Let X be a random variable and ψ_X its PGF. We have then that:*

1. *If X takes a finite number of real values then*

$$\mathbb{D}_X =]0, +\infty[\ .$$

2. *If X takes an infinite number of real values with no accumulation points then:*

$$\exists u_0, v_0 \in]-\infty, 0]\ ,\ ]e^{u_0}, e^{-v_0}[\subset \mathbb{D}_X \subset [e^{u_0}, e^{-v_0}]\ . \tag{1}$$

3. *If X is a random variable with exponentially decaying tails, that is if for some $k, c > 0$ we have that $\mathbb{P}[\mid X \mid > x] \leq ke^{-cx}$ the we also have the condition expressed by formula a* (1).

The PGF fully characterizes the law of its associated random variable. In fact, two random variables will have the same distribution if and only if the correspondent PGF coincide in a neighborhood of 1. We also have for PGF a result similar to Lévy theorem for characteristic functions. If for a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ the correspondent sequence of PGF $(\psi_n)_{n\in\mathbb{N}}$ converges to ψ_X in a neighborhood of 1 then, the sequence of random variables converges in law to X.

The PGF of a discrete random variable taking integer values is most useful for the computation of the laws of sums of independent random variables of this type, mostly because for these random variables we have $p_k = \psi_X^{(n)}(0)/n!$. For general discrete random variables (not necessarily taking integer values), the symbolic computational packages allow the same calculations which were once practically possible only for integer valued random variables. The study of PGF for real valued discrete random variables is thus fully justified.

Remark 2.1. *For a random variable having a continuous law μ_X it may happen that the set $\mathbb{D}_X := \{t > 0 : \int_{\mathbb{R}} t^x d\mu_X(x) < +\infty\}$ has a non empty interior. This is the case for Gaussian and gamma random variables. In this case we will use also the notation $\psi_X(t) = \mathbb{E}[t^X]$, for $t \in \mathbb{D}_X$.*

3 On the empirical estimator of the PGF

In this section we show how to use a sample of a random variable to estimate the PGF of this random variable. The results presented are extensions to real valued random variables of the results already known for integer valued discrete random variables (see again [13]).

Let $(X_n)_{n\in\mathbb{N}}$ be a sample of a random variable X having as probability law μ_X and PGF $\psi_X(t) = \mathbb{E}[t^X]$ defined for $t \in \mathbb{D}_X$. Let us define also the empirical PGF (EPGF) by:

$$\forall t > 0 \;\; \overline{\psi}_{X,n}(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i} \,. \tag{2}$$

As we have for all $t \in \mathbb{D}_X$ that $\mathbb{E}[\overline{\psi}_{X,n}(t)] = \psi_X(t)$ then, we have that $(\overline{\psi}_{X,n}(t))_{n\in\mathbb{N}}$ is a sequence of non biased estimators of $\psi_X(t)$. A trivial application of the strong law of large numbers shows that we have for all $t \in \mathbb{D}_X$ the strong consistency of the estimator, that is, almost surely $\lim_{n\to+\infty} \overline{\psi}_{X,n}(t) = \psi_X(t)$.

As a consequence an easy application of the central limit theorem shows that for all $t \in \mathbb{D}_X$ the sequence $(n^{1/2}(\overline{\psi}_{X,n}(t) - \psi_X(t)))_{n\geq 1}$ converges in distribution to $\mathcal{N}(0, \sqrt{\psi_X(t^2) - \psi_X(t)^2})$. We may also get a uniform law of large numbers and functional central limit theorem as we will see below.

In the proof of the next result, the uniform law of large numbers for EPGF, we follow the general idea of [10] but instead of applying Lebesgue's dominated convergence theorem we apply the inverse Fatou lemma. A result of this kind, for moment generating function,s may be found in [5].

Theorem 3.1. *Let $[a,b] \subset \mathbb{D}_X \neq \emptyset$. Then we have almost surely:*

$$\lim_{n\to+\infty} \sup_{t\in[a,b]} \left|\overline{\psi}_{X,n}(t) - \psi_X(t)\right| = 0 \ .$$

Proof. By the strong law of large numbers we have that:

$$\forall k \in \mathbb{Z} \ \ p_k = \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}} \ . \tag{3}$$

Consider now ω_0 in this set of full probability. Decomposing the sum and observing, as agreed, that $\alpha_k > 0$ for $k \geq 1$, $\alpha_k < 0$ for $k < -1$ and $\alpha_0 = 0$, we get:

$$\begin{aligned}
\sup_{t\in[a,b]} \left|\overline{\psi}_{X,n}(t,\omega_0) - \psi_X(t)\right| &\leq \sup_{t\in[a,b]} \sum_{k=-\infty}^{+\infty} t^{\alpha_k} \left|p_k - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0)\right| = \\
&= \sum_{k=1}^{+\infty} b^{\alpha_k} \left|p_k - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0)\right| + \left|p_0 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i=0\}}(\omega_0)\right| \\
&+ \sum_{k=-1}^{-\infty} a^{\alpha_k} \left|p_k - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0)\right| \ .
\end{aligned}$$

We will now show that the limit of the sums in the right side of the formula above is zero when the size of the sample n grows to infinity. We will deal only with the first sum as for the second term the conclusion follows trivially from formula 3 and for the second sum the proof is similar to one we will now present for the first sum.

For this purpose we will use the inverse Fatou lemma.

Define $f_n(k)$ and $g_n(k)$ by

$$f_n(k) := b^{\alpha_k} \left| p_k - \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \right| \leq$$
$$\leq b^{\alpha_k} \left(p_k + \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \right) =: g_n(k) \ .$$

Observe that for all $n \in \mathbb{N}$,

$$\sup_{m \geq n} \frac{1}{m} \sum_{i=1}^{m} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \leq \sup_{m \geq 1} \frac{1}{m} \sum_{i=1}^{m} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \leq$$
$$\leq \begin{cases} p_k \text{ ou} \\ (1/m_1) \sum_{i=1}^{m_1} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \ , \end{cases}$$

where m_1 is such that

$$\sup_{m \geq 1} (1/m) \sum_{i=1}^{m} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) = (1/m_1) \sum_{i=1}^{m_1} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \ .$$

With μ_c the counting measure over $\mathbb{Z}$, we have for all $n \in \mathbb{N} - \{0\}$:

$$\sum_{k=1}^{+\infty} \sup_{m \geq n} f_m(k) = \int_{\mathbb{N}-\{0\}} \sup_{m \geq n} f_m(k) d\mu_c(k) \leq \int_{\mathbb{N}-\{0\}} \sup_{m \geq n} g_m(k) d\mu_c(k) \leq$$
$$\leq \sum_{k=1}^{+\infty} b^{\alpha_k} p_k + \max \left(\sum_{k=1}^{+\infty} b^{\alpha_k} p_k, \sum_{k=1}^{+\infty} (b^{\alpha_k} \frac{1}{m_1} \sum_{i=1}^{m_1} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0)) \right) =$$

$$= \psi_X(b) + \max \left(\psi_X(b), \frac{1}{m_1} \sum_{i=1}^{m_1} b^{\alpha_{k(i,\omega_0)}} \right) < +\infty \ ,$$

noticing that for a given ω_0 and $i = 1, \ldots, m_1$ there exists only one $k = k(i, \omega_0)$ such that $\mathbb{I}_{\{X_i=\alpha_{k(i,\omega_0)}\}}(\omega_0) \neq 0$. We may now apply the inverse Fatou lemma to conclude that

$$0 \leq \limsup_{n \to +\infty} \sum_{k=1}^{+\infty} b^{\alpha_k} \left| p_k - \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \right| \leq$$
$$\leq \sum_{k=1}^{+\infty} \limsup_{n \to +\infty} \left| p_k - \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i=\alpha_k\}}(\omega_0) \right| = 0 \ ,$$

as desired $\square$

Remark 3.1. *The conclusions of this theorem remain valid under the weaker hypothesis of $(X_n)_{n\in\mathbb{N}}$ being a stationary ergodic sequence as in this case we still have formula 3.*

We also have an invariance principle for the EPGF.

Theorem 3.2. *The sequence $(n^{1/2}(\overline{\psi}_{X,n}(t)-\psi_X(t))_{n\geq 1}$ of stochastic processes converges weakly to a Gaussian process with mean zero and covariance given by $\psi_X(st)-\psi_X(s)\psi_X(t)$, on any closed interval subset of $\mathbb{D}_X/2$ and in the space of continuous functions with the uniform norm.*

Proof. This result may be deduced directly from theorem 2.3 in [5] where the result is formulated and proved for the moment generator function, $\tilde{\mu}_X(t)=\mathbb{E}[e^{tX}]$, noticing that $\psi_X(t)=\tilde{\mu}_X(\ln(t))$. □

4 Applications of the PGF

The results of the previous section allow us to consider the study of discrete real valued random variables in the same way as usually done for integer valued discrete random variables.

4.1 Application to discrete random variables

As a consequence of the results in section 3 we may apply to discrete random variables taking real values the estimation procedures developed for discrete random variables taking integer values. See for instance [11] for a complete review of these techniques and [6], [8], [5], [18], [12], [13], [2], [17] and finally [16], for detailed studies of some particular statistical studies. A detailed analysis of these methods for relevant examples of real valued discrete random variables is yet to be done.

4.2 Application to general random variables

In this subsection we show how the PGF may be used to study the law of a random variable, not necessarily discrete. The idea behind such an approximation (see [15, p.131]) is that it is physically possible to observe only a finite number (an so a discrete set) of values that a random variable takes. For this, it is appropriate to say that we may only *know* discrete random variables.

The empirical measure defined in the usual way from a sample of a random variable, see [7], approximates the law of this random variable, on every desirable aspect (uniform strong law, functional central limit theorem, etc) in a wide spectrum of situations. This empirical measure is the adequate tool for random variables taking vectorial values.

The following simple result shows that any probability law in $\mathbb{R}$ may be approximated by a family of laws of discrete random variables, built as an histogram.

It is well known that any measure in a locally compact space may be approximated by a sequence of linear combinations of Dirac measures (see [9, p. 99]).

In a parallel line of thought, a classical result shows that the histogram, built over a sample of a given random variable is a random stair function converging in probability to the density, in each continuity point of this density (see [14, p. 367]). With these results in mind it is natural to think that the law of a random variable may be approximated by a sequence of random measures built over the sample.

Theorem 4.1. *Let for each $n \in \mathbb{N}$, $(I_k^n)_{k\in\mathbb{Z}}$be a partition of the real numbers such that, $\lim_{n\to+\infty} max_{k\in\mathbb{N}} |I_k^n| = 0$ and for each $n \in \mathbb{N}$, $(\alpha_k^n)_{k\in\mathbb{Z}}$ be the sequence of left extremities of the partition intervals of order n. Let $(X_n)_{n\in\mathbb{N}}$ be a sample of a random variable X having as law μ_X. Define*

$$\overline{\mu}_{N,n} := \sum_{k\in\mathbb{Z}} \frac{\#\{i \in \{1. \dots, N\} : X_i \in I_k^n\}}{N} \delta_{\alpha_k^n} \ . \tag{4}$$

Then, $(\overline{\mu}_{N,n})_{N,n\in\mathbb{N}}$ is a family of random probability laws converging narrowly *in probability to μ_X, that is, for every f bounded and continuous*

$$\lim_{N,n\to+\infty} \overline{\mu}_{N,n}(f) = \mu_X(f) \ .$$

Proof. A simple computation shows that for every bounded and continuous f we have

$$\mathbb{E}[\overline{\mu}_{N,n}(f)] = \sum_{k\in\mathbb{Z}} \mu_X(I_k^n) f(\alpha_k^n) \ ,$$

$$\mathbb{V}[\overline{\mu}_{N,n}(f)] = \frac{1}{N} \left[\sum_{k\in\mathbb{Z}} \mu_X(I_k^n) f^2(\alpha_k^n) - \left(\sum_{k\in\mathbb{Z}} \mu_X(I_k^n) f(\alpha_k^n) \right)^2 \right] \ .$$

As a consequence, by the definition of Stieltjes integral we have that $\lim_{n\to+\infty} \mathbb{E}[\overline{\mu}_{N,n}(f)] = \mu_X(f)$ e $\lim_{n\to+\infty} N\times\mathbb{V}[\overline{\mu}_{N,n}(f)] = N\times(\mu_X(f^2) - \mu_X(f)^2)$ thus proving the result announced. □

This result will allow us to study a continuous law by means of a discrete law approximation, as justified by the next remark.

Remark 4.1. *Let $(x_1,\ldots,x_N)$ an observation of $(X_n)_{n\in\mathbb{N}}$ a sample of a random variable X having as law μ_X. Now, for n large enough and considering for a generic interval of the partition defined above $I_k^n = [x_k, x_{k+1}[$, we will have that $\#\{i \in \{1\ldots,N\} : X_i \in I_k^n\} = 1$ and so, the observation $(x_1,\ldots,x_N)$ gives us, with the notations of the theorem above, that μ is an observation of the random probability law*

$$\overline{\mu}_{N,n} = \sum_{k\in\mathbb{Z}} \frac{\#\{i \in \{1\ldots,N\} : X_i \in I_k^n\}}{N} \delta_{\alpha_k^n} = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i} \ ,$$

that converges narrowly in probability to μ_X. Being so, it is to be expected that for fixed and adequate $t \in \mathbb{D}_X$ and for all y in a compact interval we have that

$$\overline{\psi}_{X,N}(t) = \frac{1}{N}\sum_{i=1}^{N} t^{X_i} = \overline{\mu}_{N,n}(t^y) \approx \mu_X(t^y) \approx \psi_X(t) \ .$$

The first approximation being a consequence of the theorem 5 and the second deriving from X having exponentially decaying tails as in theorem 2.1 and remark 2.1.

4.3 Parameter estimation with PGF

We introduce next a PGF based estimation method for parameters of continuous random variables derived from theorem 3.1 and from remark 4.1. This is a technique usually considered useful only for discrete random variables. We will first describe the method, next we present a testing protocol for the method and finally, in section 5 we prove the consistency of the estimators in two different particular instances.

1. Consider a random variable X having the law $\mu_{X(\boldsymbol{\theta})}$ where $\boldsymbol{\theta}$ is a unknown parameter in a certain compact set $\Theta \subset \mathbb{R}^p$. Suppose that for every $\boldsymbol{\alpha} \in \Theta$ the PGF $\psi_{X(\boldsymbol{\alpha})}$ is well defined in a set $\mathbb{D}_{X(\boldsymbol{\alpha})}$ having a non empty interior.

2. Having observed a sample of X, consider $\overline{\psi}_{X,n}$ the EPGF based on the sample.

3. Consider a set of points $t_1, t_2, \ldots, t_M$ in $\bigcap_{\boldsymbol{\alpha}\in\Theta} \mathrm{Int}\mathbb{D}_{X(\boldsymbol{\alpha})}$ that we suppose to be non empty and define the contrast

$$O_n(\boldsymbol{\alpha}) := \sum_{i=1}^{M} \left(\overline{\psi}_{X,n}(t_i) - \psi_{X(\boldsymbol{\alpha})}(t_i)\right)^2 \ ,$$

and the minimum contrast estimator $\hat{\boldsymbol{\theta}}_n$ of the unknown parameter $\boldsymbol{\theta}$, such that

$$O(\hat{\boldsymbol{\theta}}_n) = \min\{O_n(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Theta\} \ .$$

In order to test this estimation procedure by simulation we propose the following protocol.

- Step 1: Choose a value for the unknown parameter $\boldsymbol{\theta}$. Let $j = 1$. Choose r the number of repetitions of the simulation.
- Step 2: Simulate a sample of $X(\boldsymbol{\theta})$.
- Step 3: Determine by the method described above $\hat{\boldsymbol{\theta}}_{1,j}$ an estimated value of $\boldsymbol{\theta}$ and, by another standard and known method, $\hat{\boldsymbol{\theta}}_{2,j}$ another estimated value for the parameter $\boldsymbol{\theta}$. If $j < r$ increment j and return to step 2. If $j = r$ go to step 4.
- Step 4: Calculate the mean and standard deviation of the families of estimated values $U = (\hat{\boldsymbol{\theta}}_{1,j})_{j=1,\ldots,r}$ and $V = (\hat{\boldsymbol{\theta}}_{2,j})_{j=1,\ldots,r}$ and compare the methods comparing the correspondent means and standard deviations.

We present next an application of this protocol for the test of the algorithm in three distinct situations in the case of a parameter of dimension one.

For the Gaussian law $\mathcal{N}(\theta, \sigma)$, that is, with mean equal to θ and standard deviation equal to σ we have that the theoretical PGF is given for $t > 0$ by $\psi_{X(\theta,\sigma)}(t) = \mathbb{E}[t^X] = t^\theta e^{\sigma \ln^2(t)/2}$. Given $\sigma = 1$, we estimate θ by the proposed method and also considering for each repetition j of the simulation, $\theta_{2,j}$ given by the average of the sample. The points $t_1, \ldots, t_M$ were chosen close to 1, more precisely, $t_1 = .8, t_2 = .85, t_3 = .9, t_4 = .95, t_5 = .98, t_6 = 1.05, t_7 = 1.08, t_8 = 1.09, t_9 = 1.1, t_{10} = 1.2$. Results for objective values $\theta = 2$, $\sigma = 1$ and for a sample of dimension 20 are shown in table 1.

Next, we consider the gamma distribution with parameters λ and α, having a density given by $\mathcal{G}(\lambda, \frac{1}{\alpha}) = \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha x} x^{\lambda-1}$. It is easy to see that if $X = X(\lambda, \alpha) \in \mathcal{G}(\lambda, 1/\alpha)$ then, for $t \in \mathbb{D}_{X(\lambda,\alpha)} =]0, e^\alpha[$ we have that $\psi_{X(\lambda,\alpha)}(t) = \mathbb{E}[t^X] = \frac{\alpha^\lambda}{(\alpha - \ln(t))^\lambda}$ As $\mathbb{E}[X(\lambda, \alpha)] = \lambda/\alpha$ e $\mathbb{V}[X(\lambda, \alpha)] = \lambda/\alpha^2$. A natural way to estimate α from a given sample consists in computing $\mathbb{E}[X(\lambda, \alpha)]/\mathbb{V}[X(\lambda, \alpha)]$. We applied the protocol defined above to estimate $\alpha = 2$ with $\lambda = 2.1$ and with the points $t_1, \ldots, t_M$ and sample

r	Average U	St. Dev. U	Average V	St. Dev. V
10	1.99842	0.246679	1.99985	0.2467220
50	1.99401	0.149717	1.99484	0.1498410
100	2.00760	0.099236	2.00748	0.0996039
500	1.99908	0.052912	1.99890	0.0520873
1000	2.00133	0.032009	2.00124	0.0319557

Table 1: Results for the Gaussian law.

r	Average U	St. Dev.U	Average V	St. Dev.V
10	2.14098	0.5723900	2.95111	1.946780
50	2.00263	0.2076540	2.12457	0.561632
100	2.02520	0.1251470	2.17005	0.334930
500	1.99878	0.0606117	2.03207	0.190385
1000	1.99331	0.0470203	1.99880	0.123071

Table 2: Results for the gamma distribution..

dimension chosen as in the previous example. The results are presented in table 2.

Finally we consider a uniform distribution on an interval $[\theta, \theta + 1]$. For this law the PGF is given by $\psi_{(\theta,1)}(t) = (t-1)t^{\theta}/\ln(t)$. The usual estimator of θ is the minimum of the sample. The objective value is $\theta = \pi$ and all the other conditions for the protocol are the same as in the two previous examples.

r	Average U	St. Dev. U	Average V	St. Dev. V
10	3.15507	0.028903	3.19542	0.0336867
50	3.13799	0.059152	3.19380	0.043032
100	3.14248	0.062015	3.18861	0.0469901
500	3.13868	0.065901	3.19004	0.0481087
1000	3.14091	0.065709	3.19062	0.0465327

Table 3: Results for the uniform law.

We may propose a preliminary conclusion. With the simulation protocol considered, the PGF estimator introduced behaves similarly as the usual estimator of the mean in the Gaussian case and has a better behavior than the moment estimator for the parameter α of the gamma law given by the ratio of the mean over the variance of the sample and also a better behavior than the minimum estimator for the θ parameter of the uniform law above.

Remark 4.2. *The set of points used to define the minimum contrast estimator will deserve some attention in future work. It is conjectured that the speed of convergence will depend on the number and distribution around 1 of these points (see the remarks in the text after theorem 2.1).*

5 On PGF based minimum contrast estimators

Under sufficiently general hypothesis it is possible to show that the minimum contrast estimator used in the examples in this work is consistent. For the reader convenience we quote here some notations and a general and useful result from [1, p. 93]) that will allow us to prove the consistency of the estimators presented above.

Let $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\boldsymbol{\theta}\in\Theta})$ be a statistical model, that is, $(\Omega, \mathcal{F})$ is a measurable space and $(\mathbb{P}_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ is a family of probability laws depending on a parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. For $\boldsymbol{\theta}_0 \in \Theta$ fixed, we consider a contrast function $K(\boldsymbol{\theta}_0, \boldsymbol{\alpha})$ to be some measurable real valued function defined for $\boldsymbol{\alpha} \in \Theta$ having a strict minimum for $\boldsymbol{\alpha} = \boldsymbol{\theta}_0$. Supposing that the experiments are described by a filtration $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$, a contrast process for $\boldsymbol{\theta}_0$ and K is a family of stochastic processes $(U_n(\boldsymbol{\alpha}))_{n\geq 0, \boldsymbol{\alpha}\in\Theta}$, independent of $\boldsymbol{\theta}_0$ such that:

- For each $\boldsymbol{\alpha} \in \Theta$ the process $(U_n(\boldsymbol{\alpha}))_{n\geq 0}$ is $\mathbb{F}$ adapted.
- For each $\boldsymbol{\alpha} \in \Theta$, $\lim_{n\to+\infty} U_n(\alpha) = K(\boldsymbol{\theta}_0, \boldsymbol{\alpha})$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$ probability.

A minimum contrast estimator associated with U is a $\mathcal{F}$ adapted estimator $(\hat{\boldsymbol{\theta}}_n)_{n\geq 1}$ such that for all $n \geq 1$ we have:

$$U_n(\boldsymbol{\theta}_n) = \inf\{U_n(\boldsymbol{\alpha})) : \boldsymbol{\alpha} \in \Theta\} .$$

With these notations and definitions we now have the following result.

Theorem 5.1 (Dacunha-Castelle & Duflo 1983). *Suppose that Θ is compact and that the real valued functions defined for $\boldsymbol{\alpha} \in \Theta$ by $K(\boldsymbol{\theta}_0, \alpha)$ and $U_n(\boldsymbol{\alpha})$ are continuous. Define for any $\eta > 0$*

$$w(n, \eta) := \sup\{|U_n(\boldsymbol{\alpha}) - U_n(\boldsymbol{\beta})| : |\boldsymbol{\alpha} - \boldsymbol{\beta}| \leq \eta\}$$

and suppose that for a sequence $(\epsilon_k)_{k\geq 1}$, decreasing to zero, we have

$$\lim_{n\to+\infty} \mathbb{P}_{\boldsymbol{\theta}_0}\left[w\left(n, \frac{1}{k}\right) \geq \epsilon_k\right] = 0 .$$

Then, any minimum contrast estimator $(\hat{\boldsymbol{\theta}}_n)_{n\geq 1}$ is consistent on $\boldsymbol{\theta}_0$.

Using this result it is now possible to show that for a class of statistical models we have consistency of the minimum contrast estimators based on PGF.

Theorem 5.2. *Let $\Theta \subset \mathbb{R}^p$ be a compact set and for every $\boldsymbol{\theta} \in \Theta$ let $f_{\boldsymbol{\theta}}$ be the density of the law $\mathbb{P}_{\boldsymbol{\theta}}$ with respect to the Lebesgue measure. We will suppose that for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \Theta$, there is some strictly positive constants $a_1, \ldots, a_N$ and some $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N \in \Theta$ such that for all $t \in \cap_{\boldsymbol{\alpha}\in\Theta}\mathbb{D}_{X(\boldsymbol{\alpha})}$ we have that, for some real function $g(t)$ not depending on $\theta \in \Theta$,*

$$\left|\psi_{X(\boldsymbol{\beta})}(t) - \psi_{X(\boldsymbol{\alpha})}(t)\right| \leq g(t)\left|\boldsymbol{\alpha} - \boldsymbol{\beta}\right| \sum_{k=1}^{N} a_k \psi_{X(\boldsymbol{\alpha}_k)}(t)\ . \tag{5}$$

Let $(X_n)_{n\in\mathbb{N}}$ be a sample of X having law $\mathbb{P}_{\boldsymbol{\theta}_0}$ and

$$\overline{\psi}_{X,n}(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i}$$

defined for $t > 0$. Define for some $M \geq 1$ and $t_1, \ldots, t_M$ in the set $\cap_{\boldsymbol{\alpha}\in\Theta}\mathbb{D}_{X(\boldsymbol{\alpha})} \neq \emptyset$, the contrast process by:

$$O_n(\boldsymbol{\theta}_0, \boldsymbol{\alpha}) := \sum_{i=1}^{M}(\overline{\psi}_{X,n}(t_i) - \psi_{X(\boldsymbol{\alpha})}(t_i))^2$$

and $(\hat{\boldsymbol{\theta}}_n)_{n\in\mathbb{N}}$ a sequence of of minimum contrast estimators of $\boldsymbol{\theta}_0$, that is, verifying for all $n \in \mathbb{N}$

$$O_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n) := \min\{O_n(\boldsymbol{\theta}_0, \boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Theta\}\ .$$

Then, $(\hat{\boldsymbol{\theta}}_n)_{n\in\mathbb{N}}$ converges in probability to $\boldsymbol{\theta}_0$.

Proof. We will apply theorem 5.1. Accordingly, we have to prove that the contrast function $K(\boldsymbol{\theta}_0, \boldsymbol{\alpha})$, which is well defined as a consequence of the law of large numbers, for instance in theorem 3.1, by:

$$K(\boldsymbol{\theta}_0, \boldsymbol{\alpha}) = \lim_{n\to+\infty} O_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n) = \sum_{i=1}^{M}(\psi_{X(\boldsymbol{\theta}_0)}(t_i) - \psi_{X(\boldsymbol{\alpha})}(t_i))^2\ ,$$

is a continuous function of the variable $\boldsymbol{\alpha}$ in Θ. This is in fact true not only for $K(\boldsymbol{\theta}, \boldsymbol{\alpha})$ but also for $O_n(\boldsymbol{\theta}, \boldsymbol{\alpha})$ by the uniform convergence, as $\psi_{X(\boldsymbol{\alpha})}(t)$, for fixed t, is a continuous function of $\boldsymbol{\alpha}$. It is clear that

$K(\boldsymbol{\theta}_0, \boldsymbol{\alpha}) \geq 0$ and that $K(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 0$. We may then conclude that for all $n \in \mathbb{N}$ the minimum contrast $\hat{\boldsymbol{\theta}}_n$ exists. Define now

$$\forall k \geq 1 \;\; w(n,k) := \sup \left\{ |O_n(\boldsymbol{\theta}_0, \boldsymbol{\alpha}) - O_n(\boldsymbol{\theta}_0, \boldsymbol{\beta})| : |\boldsymbol{\alpha} - \boldsymbol{\beta}| < \frac{1}{k} \right\} .$$

In order to have the consistency, we will verify that there exists a sequence $(\epsilon_k)_{k \geq 1}$, decreasing to zero, and such that for all $k \geq 1$ we have $\lim_{n \to +\infty} \mathbb{P}_{\boldsymbol{\theta}_0}[w(n,k) \geq \epsilon_k] = 0$. For that purpose, observe that

$$O_n(\boldsymbol{\theta}_0, \boldsymbol{\alpha}) - O_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}) =$$
$$= \sum_{i=1}^{M} [2\overline{\psi}_{X,n}(t_i) \left(\psi_{X(\boldsymbol{\beta})}(t_i) - \psi_{X(\boldsymbol{\alpha})}(t_i) \right) + \left(\psi_{X(\boldsymbol{\alpha})}(t_i)^2 - \psi_{X(\boldsymbol{\beta})}(t_i)^2 \right)]$$

and that, if we define

$$\forall k \geq 1 \;\; v(k) := \max_{1 \leq i \leq M} \sup \left\{ \left| \psi_{X(\boldsymbol{\beta})}(t_i) - \psi_{X(\boldsymbol{\alpha})}(t_i) \right| : |\boldsymbol{\alpha} - \boldsymbol{\beta}| < \frac{1}{k} \right\}$$

and

$$\forall k \geq 1 \;\; u(k) := \max_{1 \leq i \leq M} \sup \left\{ \left| \psi_{X(\boldsymbol{\beta})}(t_i)^2 - \psi_{X(\boldsymbol{\alpha})}(t_i)^2 \right| : |\boldsymbol{\alpha} - \boldsymbol{\beta}| < \frac{1}{k} \right\}$$

the sequences $(v_k)_{k \geq 1}$ and $(u_k)_{k \geq 1}$ are decreasing and so we have

$$w(n,k) \leq (u_k + v_k) \left[\sum_{i=1}^{M} (1 + 2\overline{\psi}_{X,n}(t_i)) \right] . \tag{6}$$

Considering now $w_k := u_k + v_k$, the fact that the sequence $(w_k)_{k \geq 1}$ is decreasing and the facts that $\mathbb{E}[M + 2\sum_{i=1}^{M} \overline{\psi}_{X,n}(t_i)] = M + 2\sum_{i=1}^{M} \psi_{X(\boldsymbol{\theta}_0)}(t_i)$ and also

$$\mathbb{V}[M + 2\sum_{i=1}^{M} \overline{\psi}_{X,n}(t_i)] = \frac{4}{n} \sum_{i,j=1}^{M} (\psi_{X(\boldsymbol{\theta}_0)}(t_i t_j) - \psi_{X(\boldsymbol{\theta}_0)}(t_i)\psi_{X(\boldsymbol{\theta}_0)}(t_j)) ,$$

we have the following chain of inequalities for all $c > 0$,

$$\begin{aligned}
&\mathbb{P}\left[w(n,k) \geq w_k\left((M + 2\sum_{i=1}^{M}\psi_{X(\boldsymbol{\theta}_0)}(t_i) + c\right)\right] \leq_{(a)} \\
&\leq_{(a)} \mathbb{P}\left[\sum_{i=1}^{M}\overline{\psi}_{X,n}(t_i) \geq \sum_{i=1}^{M}\psi_{X(\boldsymbol{\theta}_0)}(t_i) + \frac{c}{2}\right] \leq \\
&\leq \mathbb{P}\left[\left|(\sum_{i=1}^{M}\overline{\psi}_{X,n}(t_i)) - \sum_{i=1}^{M}\psi_{X(\boldsymbol{\theta}_0)}(t_i)\right| \geq \frac{c}{2}\right] \leq \\
&\leq \frac{4}{c^2}\mathbb{V}\left[M + 2\sum_{i=1}^{M}\overline{\psi}_{X,n}(t_i)\right] = \\
&= \frac{16}{nc^2}\sum_{i,j=1}^{M}\left(\psi_{X(\boldsymbol{\theta}_0)}(t_i t_j) - \psi_{X(\boldsymbol{\theta}_0)}(t_i)\psi_{X(\boldsymbol{\theta}_0)}(t_j)\right) \xrightarrow[n\to+\infty]{} 0 \ ,
\end{aligned}$$

where inequality (a) results from formula 6. Defining now $\epsilon_k := w_k(M + 2\sum_{i=1}^{M}\psi_{X(\boldsymbol{\theta}_0)}(t_i) + c)$ the proof will be finished as soon as we show that $\lim_{k\to+\infty} u_k = 0 = \lim_{k\to+\infty} v_k$. But this, for v_k, is a straightforward consequence of formula (5). The conclusion $\lim_{k\to+\infty} u_k = 0$ also follows as we have, as a consequence of the hypothesis given by formula (5), that choosing some fixed $\boldsymbol{\theta} \in \Theta$, we get for all $\boldsymbol{\alpha} \in \Theta$:

$$\psi_{X(\boldsymbol{\alpha})}(t) \leq \psi_{X(\boldsymbol{\theta})}(t) + g(t)\,|\boldsymbol{\alpha} - \boldsymbol{\theta}|\sum_{k=1}^{N} a_k\psi_{X(\boldsymbol{\alpha}_k)}(t) \ .$$

thus showing that for some strictly positive constants $a'_1, \ldots, a'_{N+1}$ and some $\boldsymbol{\alpha}'_1, \ldots, \boldsymbol{\alpha}'_{N+1} \in \Theta$ we have for all $t \in \cap_{\boldsymbol{\alpha}\in\Theta}\mathbb{D}_{X(\boldsymbol{\alpha})}$ and with $h(t) = 2\max(g(t), 1)$:

$$\left|\psi_{X(\boldsymbol{\beta})}(t)^2 - \psi_{X(\boldsymbol{\alpha})}(t)^2\right| \leq h(t)\left(\sum_{k=1}^{N+1} a'_k\psi_{X(\boldsymbol{\alpha}'_k)}(t)\right)\left|\psi_{X(\boldsymbol{\beta})}(t) - \psi_{X(\boldsymbol{\alpha})}(t)\right|$$

allowing us to apply the same reasoning as we did for v_k. We must remark that only a finite number of points $t \in \cap_{\boldsymbol{\alpha}\in\Theta}\mathbb{D}_{X(\boldsymbol{\alpha})}$ intervene in the definitions of v_k e u_k. □

5.1 Examples

We now present applications of the method introduced in this work to the estimation of parameters of some distributions.

Theorem 5.3 (PGF estimation of the mean a Gaussian distribution)**.** *Let $X = X(\theta) \in \mathcal{N}(\theta, \sigma)$ such that σ is given and where $\theta \in \Theta \subset \mathbb{R}$ with Θ a compact set . We then have that for all $\alpha, \beta \in [\theta_-, \theta_+]$ and for any small $\epsilon > 0$ that, with $\theta_* := \min(|\theta^+|, |\theta^-|)$,*

$$
\begin{aligned}
&\left|\psi_{X(\beta)}(t) - \psi_{X(\alpha)}(t)\right| \leq \\
&\leq \frac{|\alpha-\beta|}{\sigma}\left((|\alpha| + |\beta|)\left[e^{\frac{\theta_+^2-\theta_*^2}{2\sigma}}\psi_{X(\theta_+)}(t) + e^{\frac{\theta_-^2-\theta_*^2}{2\sigma}}\psi_{X(\theta_-)}(t)\right] + \right. \\
&\left. \frac{\sigma}{\epsilon e}\left[e^{\frac{\epsilon^2+2\epsilon\theta_+}{2\sigma}+\frac{\theta_+^2-\theta_*^2}{2\sigma}}\psi_{X(\theta_++\epsilon)}(t) + e^{\frac{\epsilon^2-2\epsilon\theta_-}{2\sigma}+\frac{\theta_-^2-\theta_*^2}{2\sigma}}\psi_{X(\theta_--\epsilon)}(t)\right]\right) .
\end{aligned}
\tag{7}
$$

And so, the PGF based estimator of the mean of a Gaussian random variable given by theorem 5.2 is consistent.

Proof. Let f_θ denote the density of $X = X(\theta) \in \mathcal{N}(\theta, \sigma)$. We have that by the mean value theorem:

$$
\begin{aligned}
\psi_{X(\beta)}(t) - \psi_{X(\alpha)}(t) &= \int_{\mathbb{R}} t^x (f_\beta(x) - f_\alpha(x))dx = \\
&= \int_{\mathbb{R}} t^x (\beta - \alpha)\left[\frac{\partial}{\partial\theta} f_\theta(x)\right]_{\theta_x := \lambda_x\alpha + (1-\lambda_x)\beta} dx .
\end{aligned}
\tag{8}
$$

As a consequence we will have that

$$
\left|\psi_{X(\beta)}(t) - \psi_{X(\alpha)}(t)\right| \leq \frac{|\alpha-\beta|}{\sigma\sqrt{2\pi\sigma}} \int_{\mathbb{R}} t^x (|x| + |\alpha| + |\beta|) e^{\frac{-(x-\theta_x)^2}{2\sigma}} dx \tag{9}
$$

The proof of inequality (7) will be completed by the estimation of two integrals. As for the simplest one, we have that:

$$
\begin{aligned}
&\int_{\mathbb{R}} t^x e^{\frac{-(x-\theta_x)^2}{2\sigma}} dx \leq e^{-\frac{\theta_*^2}{2\sigma}}\left(\int_{\mathbb{R}} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_x}{2\sigma}} dx\right) \leq \\
&\leq e^{-\frac{\theta_*^2}{2\sigma}}\left(\int_{\mathbb{R}_+} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_+}{2\sigma}} dx + \int_{\mathbb{R}_-} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_-}{2\sigma}} dx\right) = \\
&= e^{-\frac{\theta_*^2}{2\sigma}}\left(e^{+\frac{\theta_+^2}{2\sigma}}\int_{\mathbb{R}_+} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_+}{2\sigma}-\frac{\theta_+^2}{2\sigma}} dx + \right. \\
&\left. + e^{+\frac{\theta_-^2}{2\sigma}}\int_{\mathbb{R}_-} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_-}{2\sigma}-\frac{\theta_-^2}{2\sigma}} dx\right) = \\
&= e^{\frac{\theta_+^2-\theta_*^2}{2\sigma}}\int_{\mathbb{R}_+} t^x e^{-\frac{(x-\theta_+)^2}{2\sigma}} dx + e^{\frac{\theta_-^2-\theta_*^2}{2\sigma}}\int_{\mathbb{R}_-} t^x e^{-\frac{(x-\theta_-)^2}{2\sigma}} dx \leq \\
&\leq \sqrt{2\pi\sigma}\left(e^{\frac{\theta_+^2-\theta_*^2}{2\sigma}}\psi_{X(\theta_+)}(t) + e^{\frac{\theta_+^2-\theta_*^2}{2\sigma}}\psi_{X(\theta_-)}(t)\right)
\end{aligned}
$$

We now deal with the second integral. For that we will compensate the additional factor $|x|$ by some exponential term. Starting as in the first integral we will have to evaluate an integral of the form:

$$\int_{\mathbb{R}_+} t^x x e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_+}{2\sigma}-\frac{\theta_+^2}{2\sigma}} dx =$$
$$= \int_{\mathbb{R}_+} t^x \left(xe^{-\frac{2x\epsilon}{2\sigma}}\right) e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_+}{2\sigma}-\frac{\theta_+^2}{2\sigma}+\frac{2x\epsilon}{2\sigma}} dx \leq$$
$$\leq \frac{\sigma}{\epsilon e} \int_{\mathbb{R}_+} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x(\theta_+ + \epsilon)}{2\sigma}-\frac{\theta_+^2}{2\sigma}} dx =$$
$$= \frac{\sigma}{\epsilon e} e^{\frac{\epsilon^2+2\epsilon\theta_+}{2\sigma}} \int_{\mathbb{R}_+} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x}{2\sigma}(\theta_+ + \epsilon)-\frac{(\theta_+ + \epsilon)^2}{2\sigma}} dx =$$
$$= \frac{\sigma}{\epsilon e} e^{\frac{\epsilon^2+2\epsilon\theta_+}{2\sigma}} \int_{\mathbb{R}_+} t^x e^{-\frac{(x-(\theta_+ + \epsilon))^2}{2\sigma}} dx \leq \sqrt{2\pi\sigma}\frac{\sigma}{\epsilon e} e^{\frac{\epsilon^2+2\epsilon\theta_+}{2\sigma}} \psi_{X(\theta_+ + \epsilon)}(t)$$

With the same reasoning:

$$\int_{\mathbb{R}_-} t^x (-x) e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_-}{2\sigma}-\frac{\theta_-^2}{2\sigma}} dx =$$
$$= \int_{\mathbb{R}_-} t^x \left(-xe^{\frac{2x\epsilon}{2\sigma}}\right) e^{-\frac{x^2}{2\sigma}+\frac{2x\theta_-}{2\sigma}-\frac{\theta_-^2}{2\sigma}-\frac{2x\epsilon}{2\sigma}} dx \leq$$
$$\leq \frac{\sigma}{\epsilon e} \int_{\mathbb{R}_-} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x(\theta_- - \epsilon)}{2\sigma}-\frac{\theta_-^2}{2\sigma}} dx =$$
$$= \frac{\sigma}{\epsilon e} e^{\frac{\epsilon^2-2\epsilon\theta_-}{2\sigma}} \int_{\mathbb{R}_-} t^x e^{-\frac{x^2}{2\sigma}+\frac{2x}{2\sigma}(\theta_- - \epsilon)-\frac{(\theta_- - \epsilon)^2}{2\sigma}} dx =$$
$$= \frac{\sigma}{\epsilon e} e^{\frac{\epsilon^2-2\epsilon\theta_-}{2\sigma}} \int_{\mathbb{R}_-} t^x e^{-\frac{(x-(\theta_- - \epsilon))^2}{2\sigma}} dx \leq \sqrt{2\pi\sigma}\frac{\sigma}{\epsilon e} e^{\frac{\epsilon^2-2\epsilon\theta_-}{2\sigma}} \psi_{X(\theta_- - \epsilon)}(t)$$

Formula 7 shows that the hypothesis of theorem 5.2, namely formula (5) are verified. The result now follows. □

We will now deal with the gamma distribution with shape parameter λ and rate parameter θ, that is an element of $\mathcal{G}(\lambda, \theta)$. Such a distribution has a density given by

$$f_{(\lambda,'\theta)}(x) = \frac{\theta^\lambda}{\Gamma(\lambda)} e^{-\theta x} x^{\lambda-1} \mathbb{I}_{]0,+\infty[}$$

See also section 4.3 for further notation.

Theorem 5.4 (PGF estimation of the rate parameter of a gamma distribution). *Let $X = X(\lambda,\theta) \in \mathcal{G}(\lambda,\theta)$ such that the shape parameter λ is given and the unknown parameter verifies $\theta \in \Theta = [\theta_-, \theta_+] \subset \mathbb{R}$, with $[\theta_-, \theta_+]$ a compact interval. We have that for all $\alpha, \beta \in [\theta_-, \theta_+]$ and $t \in]0, e^{\theta_-}[$:*

$$\begin{aligned}\left|\psi_{X(\lambda,\beta)}(t) - \psi_{X(\lambda,\alpha)}(t)\right| \leq |\alpha - \beta| \Bigg(& \frac{\lambda \theta_+^{\lambda-1}}{\theta_-^{\lambda}} \psi_{X(\theta_-,\lambda)}(t) + \\ & + \frac{\lambda \theta_+^{\lambda}}{\theta_-^{\lambda+1}} \psi_{X(\theta_-,\lambda+1)}(t) \Bigg) ,\end{aligned} \tag{10}$$

Then, the PGF minimum contrast estimator of the parameter θ given by theorem 5.2 is consistent.

Proof. The proof of inequality (10) goes exactly as the proof of the correspondent inequality for the Gaussian distribution above, although in a simpler manner. □

For the uniform distribution the result is even simpler.

Theorem 5.5 (PGF estimation of the lower parameter of a uniform distribution). *Let $X = X(\theta,\sigma) \in \mathcal{U}(\theta,\sigma)$ such that the σ is given and the unknown parameter θ verifies $\theta \in \Theta = [\theta_-, \theta_+] \subset \mathbb{R}$, with $[\theta_-, \theta_+]$ a compact interval. We have that for all $\alpha, \beta \in [\theta_-, \theta_+]$ and $t > 0$:*

$$\left|\psi_{X(\alpha,1)}(t) - \psi_{X(\beta,1)}(t)\right| \leq |\ln(t)| \, |\alpha - \beta| \left(\psi_{X(\theta_+,1)}(t) + \psi_{X(\theta_-,1)}(t)\right) \tag{11}$$

Then, the PGF minimum contrast estimator of the parameter θ given by theorem 5.2 is consistent.

Proof. The proof of formula 11 goes along similar lines as the previous proofs in this section. In fact, with the computation of the PGF of the uniform law mentioned above, we have that:

$$\left|\psi_{X(\alpha,1)}(t) - \psi_{X(\beta,1)}(t)\right| = \left|\frac{t-1}{\ln(t)}\right| \left|t^{\alpha} - t^{\beta}\right| = \left|\frac{t-1}{\ln(t)}\right| \left|\int_{\alpha}^{\beta} \ln(t) t^{x} dx\right|$$

As we now have that:

$$\begin{aligned}\sup_{x \in [\alpha,\beta]} t^x \leq \sup_{x \in [\theta_-,\theta_+]} t^x \leq & \begin{cases} t^{\theta_+} & \text{for } t \geq 1 \\ t^{\theta_-} & \text{for } t \leq 1 \end{cases} \\ & = t^{\theta_+} \mathbb{I}_{\{t \geq 1\}} + t^{\theta_-} \mathbb{I}_{\{t \leq 1\}} ,\end{aligned}$$

the result follows at once. □

Thanks

This work was partially supported by *Financiamento Base 2008 ISFL-1-297* from FCT/MCTES/PT.

References

[1] D. Dacunha-Castelle, M. Duflo (1983). *Probabilités et Statistiques*, 2. Problèmes à temps mobile. Masson, Paris.

[2] M. M. Dowling, M. Nakamura (1997). Estimating parameters for discrete distributions via the empirical probability generating function. *Commun. Stat., Simulation Comput. 26* **1**, p. 301–313.

[3] Esquível, M. L. (2007) Probability generating functions for discrete real-valued random variables. *Teor. Veroyatn. Primen.* **52** , no. 1, 129–149; translation in *Theory Probab. Appl.* **52** (2008), no. 1, 40–57.

[4] Esquível, M. L. (2005) Aplicações das funções geradoras de probabilidade a variáveis aleatórias reais. *Proceedings of the XII Annual Congress Portuguese Statistical Society.* Editors Carlos Braumann, Paulo Infante, Manuela Oliveira, Russell Alpizar Jara e Fernando Rosado, 235–246.

[5] A. Feuerverger (1989). On the empirical saddlepoint approximation, *Biometrika 3* **76**, p. 457–464.

[6] A. Feuerverger, P. McDunnough (1984). On statistical transform methods and their efficiency, *The Can. J. of Stat. 12* **4**, p. 303–317.

[7] P. Gaenssler (1983). *Empirical Processes*, Lecture Notes-Monograph Series volume 3, Institute of Mathematical Statistics.

[8] S. Kocherlakota, K. Kocherlakota (1986). Goodness of fit tests for discrete distributions, *Commun. Statist.-Theor. Meth. 15* **3**, p. 815–829.

[9] P. Malliavin (1995). *Integration and Probability*, Springer Verlag.

[10] M. S. Marques, V. Pérez-Abreu (1989). Law of large numbers and central limit theorem for the empirical probability generating function of stationary random sequences and processes. *Aportaciones Mat., Notas Invest.* 2 **4**, p. 100–109.

[11] M. Nakamura, V. Pérez-Abreu (1993). Empirical probability generating function. An overview. *Insur. Math. Econ. 12* **3**, p. 349–366.

[12] M. Nakamura, V. Pérez-Abreu (1993). Exploratory data analysis for counts using the empirical probability generating function. *Commun. Stat., Theory Methods 22* **3**, p. 827-842.

[13] M. Nakamura, V. Pérez-Abreu (1993). Use of an empirical probability generating function for testing a Poisson model. *Can. J. Stat. 21, No.2* **2**, p. 149–156.

[14] W. R. Pestman (1998). *Mathematical Statistics*, Walter de Gruyter, Paris, New York.

[15] B. L. S. Prakasa Rao (1999). *Statistical Inference for Diffusion Type Processes*, Arnold Hodder Headline Group.

[16] B. Rémillard, R. Theodorescu (2000). Inference based on the empirical probability generating function for mixtures of Poisson distributions. *Stat. Decis. 18*, **4**.

[17] R. Rueda, F. O'Reilly (1999). Tests of fit for discrete distributions based on the probability generating function. *Commun. Stat., Simulation Comput. 28* **1**, p. 259–274.

[18] R. Rueda, V. Pérez-Abreu, F. O'Reilly (1991), Goodness of fit for the Poisson distribution based on the probability generating function. *Commun. Stat., Theory Methods 20* **10**, p. 3093–3110.