1. Introduction and some structural properties of empirical measures.

Many standard procedures in statistics are based on a random sample $x_1,\dots,x_n$ of i.i.d. observations, i.e., it is assumed that observations (or measurements) occur as realizations (or values) $x_i = \xi_i(\omega)$ in some sample space X of a sequence of independent and identically distributed (i.i.d.) random elements $\xi_1,\dots,\xi_n$ defined on some basic probability space (p-space for short) $(\Omega,\mathcal{F},\mathbb{P})$; here ξ is called a RANDOM ELEMENT in X whenever there exists a $(\Omega,\mathcal{F},\mathbb{P})$ such that $\xi:\Omega\to X$ is $\mathcal{F},\mathcal{B}$-measurable for an appropriate σ-algebra $\mathcal{B}$ in X, in which case the law $\mu\equiv\mathcal{L}\{\xi\}$ of ξ is a well defined p-measure on $\mathcal{B}$ ($\mu(B)=\mathbb{P}(\{\omega\in\Omega:\ \xi(\omega)\in B\})\equiv\mathbb{P}(\xi\in B)$ for short, $B\in\mathcal{B}$).

In classical situations, the sample space X is usually the k-dimensional Euclidean space $\mathbb{R}^k$, $k\geq 1$, with the Borel σ-algebra $\mathcal{B}_k$. In the present notes, if not stated otherwise, the sample space X is always an arbitrary measurable space $(X,\mathcal{B})$.

Given then i.i.d. random elements ξ_i in $X=(X,\mathcal{B})$ with (common) law μ on $\mathcal{B}$ we can associate with each (sample size) n the so-called EMPIRICAL MEASURE

$$\mu_n := \frac{1}{n}(\varepsilon_{\xi_1}+\dots+\varepsilon_{\xi_n}) \text{ on } \mathcal{B}, \tag{1}$$

where

$$\varepsilon_x(B) := \begin{cases} 1 \text{ if } x\in B \\ 0 \text{ if } x\notin B \end{cases},\quad B\in\mathcal{B}.$$

In other words, given the first n observations $x_i=\xi_i(\omega)$, $i=1,\dots,n$, $\mu_n(B)\equiv\mu_n(B,\omega)$ is the average number of the first n x_i's falling into B. (The notation $\mu_n(\cdot,\omega)$ should call attention to the fact that μ_n is a random p-measure on $\mathcal{B}$.)

μ_n may be viewed as the statistical picture of μ and we are thus interested in the connection between μ_n and μ, especially when n tends to infinity.

In what follows, let $\mathcal{C}$ be some subset of $\mathcal{B}$ (e.g., $\mathcal{C} = \{(-\infty,t]: t\in\mathbb{R}^k\}$, the class of all lower left orthants in $X = \mathbb{R}^k$, or the class of all closed Euclidean balls in $\mathbb{R}^k$, to have at least two specific examples in mind). Denoting with 1_C the indicator function of a set $C\in\mathcal{C}$, $\mu_n(C)$ can be rewritten in the form

$$\mu_n(C) = \frac{1}{n} \sum_{i=1}^{n} 1_C(\xi_i).$$

Now, since the random variables $1_C(\xi_i)$, $i=1,2,\ldots$ are again i.i.d. with common mean $\mu(C)$ and variance $\mu(C)(1-\mu(C))$, it results from classical probability theory that

(2) (Strong Law of Large Numbers): For each fixed $C\in\mathcal{C}$ one has
$\mu_n(C) \to \mu(C)$ $\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.)
as n tends to infinity.

(3) (Central Limit Theorem): For each fixed $C\in\mathcal{C}$ one has
$n^{1/2}(\mu_n(C) - \mu(C)) \xrightarrow{\mathcal{L}} \overline{G}_\mu(C)$ as n tends to infinity,
where $\overline{G}_\mu(C)$ is a random variable with
$\mathcal{L}\{\overline{G}_\mu(C)\} = \mathcal{N}(0,\mu(C)(1-\mu(C)))$.

(4) (Multidimensional Central Limit Theorem): For any finitely many $C_1,\ldots,C_k\in\mathcal{C}$ one has
$\{n^{1/2}(\mu_n(C_j) - \mu(C_j)) : j=1,\ldots,k\} \xrightarrow{\mathcal{L}} \{\overline{G}_\mu(C_j) : j=1,\ldots,k\}$
as n tends to infinity, where $\overline{\mathbb{G}}_\mu \equiv (\overline{G}_\mu(C))_{C\in\mathcal{C}}$ is a mean-zero Gaussian process with covariance structure
$\text{cov}(\overline{G}_\mu(C), \overline{G}_\mu(D)) = \mu(C\cap D) - \mu(C)\mu(D)$, $C,D\in\mathcal{C}$.

Here, according to Kolmogorov's theorem (cf. Gaenssler-Stute (1977), 7.1.16), $\overline{\mathbb{G}}_\mu$ is viewed as a random element in $(\mathbb{R}^{\mathcal{C}},\mathbb{B}_{\mathcal{C}})$, where $\mathbb{B}_{\mathcal{C}} \equiv \bigotimes_{\mathcal{C}} \mathbb{B}$ denotes the product σ-algebra in $\mathbb{R}^{\mathcal{C}}$ of identical components $\mathbb{B}$, $\mathbb{B}$ being the σ-algebra of Borel sets in $\mathbb{R}$.

In this lecture we are going to present uniform analogues of (2) (with the uniformity being in $C\in\mathcal{C}$) known as GLIVENKO-CANTELLI THEOREMS (Section 2) and functional versions of (4), so-called FUNCTIONAL CENTRAL LIMIT THEOREMS (Section 4); an appropriate setting for the latter is presented in Section 3 which might also be of independent interest. First we want to give insight into some more or less known

STRUCTURAL PROPERTIES OF EMPIRICAL MEASURES:

For this, consider instead of μ_n the counting process

$$N_n(B) := n\mu_n(B), \quad B \in \mathcal{B}.$$

Note that $\mathcal{L}\{N_n(B)\} = \mathrm{Bin}(n,\mu(B))$ (i.e., $\mathbb{P}(N_n(B)=j) = \binom{n}{j}\mu(B)^j(1-\mu(B))^{n-j}$, $j=0,1,\ldots,n$). The following Markov and Martingale properties associated with empirical measures are well known; since however specific references are not conveniently available, and especially not in the set-indexed context of these lectures, we present detailed derivatives.

LEMMA 1 (MARKOV PROPERTY). For any $\emptyset = B_o \subset B_1 \subset \ldots \subset B_{k-1} \subset B_k \subset B_{k+1} = X$ with $B_i \in \mathcal{B}$ such that for $D_i := B_i \setminus B_{i-1}$, $\mu(D_i) > 0$, $i=1,\ldots,k+1$, and for any $0 \leqq m_1 \geqq \ldots \leqq m_{k-1} \leqq m_k \leqq n$ with $m_i \in \{0,1,\ldots,n\}$ one has

$$\mathbb{P}(N_n(B_k) = m_k \mid N_n(B_1) = m_1,\ldots,N_n(B_{k-1}) = m_{k-1})$$
$$= \mathbb{P}(N_n(B_k) = m_k \mid N_n(B_{k-1}) = m_{k-1})$$
$$= \binom{n-m_{k-1}}{m_k-m_{k-1}} \cdot \left(\frac{\mu(D_k)}{\mu(D_k \cup D_{k+1})}\right)^{m_k-m_{k-1}} \cdot \left(1 - \frac{\mu(D_k)}{\mu(D_k \cup D_{k+1})}\right)^{n-m_k}.$$

Proof. $\mathbb{P}(N_n(B_k) = m_k \mid N_n(B_1) = m_1,\ldots,N_n(B_{k-1}) = m_{k-1})$

$$= \frac{\mathbb{P}(N_n(B_1)=m_1,\ldots,N_n(B_{k-1})=m_{k-1},N_n(B_k)=m_k)}{\mathbb{P}(N_n(B_1)=m_1,\ldots,N_n(B_{k-1})=m_{k-1})} =: \frac{a}{b}, \text{ say, where}$$

$$a = \mathbb{P}(N_n(D_1) = m_1,\ N_n(D_2) = m_2-m_1,\ldots,N_n(D_k) = m_k-m_{k-1},\ N_n(D_{k+1}) = n-m_k)$$

$$= \frac{n!}{m_1(m_2-m_1)!\ldots(m_k-m_{k-1})!(n-m_k)!}\,\mu(D_1)^{m_1}\mu(D_2)^{m_2-m_1}\ldots\mu(D_k)^{m_k-m_{k-1}}\mu(D_{k+1})^{n-m_k},$$

$$b = \frac{n!}{m_1!(m_2-m_1)!\ldots(m_{k-1}-m_{k-2})!(n-m_{k-1})!}\,\mu(D_1)^{m_1}\mu(D_2)^{m_2-m_1}\ldots\mu(D_{k-1})^{m_{k-1}-m_{k-2}} \times$$
$$\times\, \mu(D_k \cup D_{k+1})^{n-m_{k-1}},$$

whence $$\frac{a}{b} = \frac{(n-m_{k-1})!}{(m_k-m_{k-1})!(n-m_k)!} \cdot \frac{\mu(D_k)^{m_k-m_{k-1}}\mu(D_{k+1})^{n-m_k}}{\mu(D_k \cup D_{k+1})^{n-m_{k-1}}}$$

$$= \binom{n-m_{k-1}}{m_k - m_{k-1}} \cdot \left(\frac{\mu(D_k)}{\mu(D_k \cup D_{k+1})} \right)^{m_k - m_{k-1}} \cdot \left(1 - \frac{\mu(D_k)}{\mu(D_k \cup D_{k+1})} \right)^{n-m_k}$$ proving equality of

the first and third term in the assertion of the lemma; the other equality follows in the same way by just taking $B_1, \ldots, B_{k-2}$ out of consideration. □

Corollary. Let $\mathcal{C}$ be a subset of $\mathcal{B}$ which is linearly ordered by inclusion; then $(N_n(C))_{C\in\mathcal{C}}$ is a Markov process.

Lemma 2. Let $\bar{B}\in\mathcal{B}$ be arbitrary but fixed such that $0 < \mu(\bar{B}) < 1$ and let $\mathcal{C} \equiv \mathcal{B}(\bar{B}) \subset \mathcal{B}$ be linearly ordered by inclusion with $\bar{B}$ as its smallest element; then for $0 \leq m \leq n$

$$\mathcal{L}\{(N_n(B))_{B\in\mathcal{B}(\bar{B})} \mid N_n(\bar{B}) = m\} = \mathcal{L}\{(m + \bar{N}_{n-m}(B\setminus\bar{B}))_{B\in\mathcal{B}(\bar{B})}\},$$

where $\bar{N}_N(D) := N\,\bar{\mu}_N(D)$, $\bar{\mu}_N$ being the empirical measure pertaining to i.i.d. random elements $\bar{\xi}_i$ in $(X,\mathcal{B})$ with $\mathcal{L}\{\bar{\xi}_i\} = \bar{\mu}$ and $\bar{\mu}(D) := \frac{\mu(D\cap \complement\bar{B})}{\mu(\complement\bar{B})}$ for $D\in\mathcal{B}$.

(Here the laws $\mathcal{L}\{\ldots\}$ are considered to be defined on the product σ-algebra $\boldsymbol{\mathcal{B}}_{\mathcal{B}(\bar{B})}$ in $\mathbb{R}^{\mathcal{B}(\bar{B})}$ and $\complement\bar{B}$ denotes the complement of $\bar{B}$ in X.)

Proof. It suffices to show that the finite dimensional marginal distributions coincide.

1) As to the one-dimensional marginal distributions, let $B\in\mathcal{B}$ with $\bar{B}\subset B$ be arbitrary but fixed; then it follows from Lemma 1 that for $k \geq m$

$$\mathbb{P}(N_n(B) = k \mid N_n(\bar{B}) = m) = \binom{n-m}{k-m} \cdot \left(\frac{\mu(B\setminus\bar{B})}{\mu(\complement\bar{B})} \right)^{k-m} \cdot \left(1 - \frac{\mu(B\setminus\bar{B})}{\mu(\complement\bar{B})} \right)^{n-k}$$

$$= \binom{n-m}{k-m} \bar{\mu}(B)^{k-m}(1 - \bar{\mu}(B))^{n-k}.$$

On the other hand, taking into account that $1_B(\bar{\xi}_i) \overset{\mathcal{L}}{=} 1_{B\setminus\bar{B}}(\bar{\xi}_i)$ (where $\overset{\mathcal{L}}{=}$ means equality in law) and therefore $\bar{N}_N(B) \overset{\mathcal{L}}{=} \bar{N}_N(B\setminus\bar{B})$ for any $B\in\mathcal{B}$ with $\bar{B}\subset B$, one obtains that

$$\binom{n-m}{k-m} \bar{\mu}(B)^{k-m}(1-\bar{\mu}(B))^{n-k} = \mathbb{P}(\bar{N}_{n-m}(B) = k-m)$$

$$= \mathbb{P}(\bar{N}_{n-m}(B\setminus\bar{B}) = k-m) = \mathbb{P}(m+\bar{N}_{n-m}(B\setminus\bar{B}) = k)$$

proving the coincidence of the one-dimensional marginal distributions.

2) As to higher-dimensional marginal distributions, let us consider for simplicity the two-dimensional case (the general case runs in the same way): For this, let $B_i \in \mathcal{B}$, i=1,2, with $\bar{B} \subset B_1 \subset B_2$ be arbitrary but fixed; then for $k_2 \geq k_1 \geq m$,

$$\mathbb{P}(N_n(B_1) = k_1,\ N_n(B_2) = k_2 \mid N_n(\bar{B}) = m)$$

$$= \frac{\mathbb{P}(N_n(\bar{B}) = m,\ N_n(B_1) = k_1,\ N_n(B_2) = k_2)}{\mathbb{P}(N_n(\bar{B}) = m)} =: \frac{a}{b}, \text{ say, where}$$

$$a = \mathbb{P}(N_n(\bar{B}) = m,\ N_n(B_1\setminus\bar{B}) = k_1-m,\ N_n(B_2\setminus B_1) = k_2-k_1,\ N_n(X\setminus B_2) = n-k_2)$$

$$= \frac{n!}{m!(k_1-m)!(k_2-k_1)!(n-k_2)!}\ \mu(\bar{B})^m \mu(B_1\setminus\bar{B})^{k_1-m}\mu(B_2\setminus B_1)^{k_2-k_1}\mu(X\setminus B_2)^{n-k_2} \quad \text{and}$$

$$b = \binom{n}{m}\mu(\bar{B})^m(1-\mu(\bar{B}))^{n-m}, \text{ whence}$$

$$\frac{a}{b} = \frac{(n-m)!}{(k_1-m)!(k_2-k_1)!(n-k_2)!}\ \frac{\mu(B_1\setminus\bar{B})^{k_1-m}\mu(B_2\setminus B_1)^{k_2-k_1}\mu(X\setminus B_2)^{n-k_2}}{\mu(\complement\bar{B})^{n-m}}$$

$$= \frac{(n-m)!}{(k_1-m)!((k_2-m)-(k_1-m))!((n-m)-(k_2-m))!} \times$$

$$\times \left(\frac{\mu(B_1\cap\complement\bar{B})}{\mu(\complement\bar{B})}\right)^{k_1-m} \cdot \left(\frac{\mu(B_2\cap\complement\bar{B})}{\mu(\complement\bar{B})} - \frac{\mu(B_1\cap\complement\bar{B})}{\mu(\complement\bar{B})}\right)^{k_2-m-(k_1-m)} \cdot \left(1 - \frac{\mu(B_2\cap\complement\bar{B})}{\mu(\complement\bar{B})}\right)^{n-m-(k_2-m)}$$

$$= \frac{(n-m)!}{(k_1-m)!((k_2-m)-(k_1-m))!((n-m)-(k_2-m))!} \times$$

$$\times\ \bar{\mu}(B_1)^{k_1-m}\bar{\mu}(B_2\setminus B_1)^{k_2-m-(k_1-m)}\bar{\mu}(\complement B_2)^{n-m-(k_2-m)}$$

$$= \frac{(n-m)!}{(k_1-m)!((k_2-m)-(k_1-m))!((n-m)-(k_2-m))!} \times$$

$$\times\ \bar{\mu}(B_1\setminus\bar{B})^{k_1-m}\bar{\mu}((B_2\setminus\bar{B})\setminus(B_1\setminus\bar{B}))^{k_2-m-(k_1-m)}\bar{\mu}(\complement(B_2\setminus\bar{B}))^{n-m-(k_2-m)}$$

$$= \mathbb{P}(\bar{N}_{n-m}(B_1\setminus\bar{B}) = k_1-m,\ \bar{N}_{n-m}(B_2\setminus\bar{B}) = k_2-m)$$

$$= \mathbb{P}(m+\bar{N}_{n-m}(B_1\setminus\bar{B}) = k_1,\ m+\bar{N}_{n-m}(B_2\setminus\bar{B}) = k_2). \quad \square$$

<u>LEMMA 3 (MARTINGALE PROPERTY).</u> Let $\mathcal{C}\subset\mathcal{B}$ be linearly ordered by inclusion such that $\mu(\complement B)>0$ for all $B\in\mathcal{C}$; then, for each fixed n,

$$\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\right)_{B\in\mathcal{C}}$$ is a martingale, i.e., for each $\bar{B}$, $B\in\mathcal{C}$ with $\bar{B}\subset B$ one has

$$\mathbb{E}\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\,\Big|\,N_n(D)\,:\,\mathcal{C}\ni D\subset\bar{B}\right) \underset{\mathbb{P}\text{-a.s.}}{=} \frac{N_n(\bar{B})-n\mu(\bar{B})}{\mu(\complement\bar{B})}\,.$$

<u>Proof.</u> Since $(N_n(C))_{C\in\mathcal{C}}$ is a Markov process (cf. Corollary to Lemma 1), it follows that

$$\mathbb{E}\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\,\Big|\,N_n(D)\,:\,\mathcal{C}\ni D\subset\bar{B}\right) \underset{\mathbb{P}\text{-a.s.}}{=} \mathbb{E}\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\,\Big|\,N_n(\bar{B})\right),$$

where $$\mathbb{E}\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\,\Big|\,N_n(\bar{B})\right)(\omega) = \mathbb{E}\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\,\Big|\,N_n(\bar{B})=m\right)$$

for all $\omega\in\{N_n(\bar{B})=m\}$, $m=0,1,\ldots,n$.

Now, according to Lemma 2, $\mathbb{E}\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\,\Big|\,N_n(\bar{B})=m\right)$

$$= \mathbb{E}\left(\frac{m+\bar{N}_{n-m}(B\setminus\bar{B})}{\mu(\complement B)}\right) - \frac{n\mu(B)}{\mu(\complement B)} = \frac{m+(n-m)\bar{\mu}(B)}{\mu(\complement B)} - \frac{n\mu(B)}{\mu(\complement B)}$$

$$= \frac{m\mu(\complement\bar{B}) + (n-m)\mu(B\cap\complement\bar{B}) - n\mu(B)\mu(\complement\bar{B})}{\mu(\complement B)\mu(\complement\bar{B})} \underset{(\mu(B\cap\complement\bar{B})=\mu(B)-\mu(\bar{B}))}{=}$$

$$= \frac{m-m\mu(\bar{B})+n\mu(B)-m\mu(B)-n\mu(\bar{B})+m\mu(\bar{B})-n\mu(B)+n\mu(B)\mu(\bar{B})}{\mu(\complement B)\mu(\complement\bar{B})}$$

$$= \frac{(1-\mu(B))(m-n\mu(\bar{B}))}{\mu(\complement B)\mu(\complement\bar{B})} = \frac{m-n\mu(\bar{B})}{\mu(\complement\bar{B})}\;;$$

hence $$\mathbb{E}\left(\frac{N_n(B)-n\mu(B)}{\mu(\complement B)}\,\Big|\,N_n(\bar{B})\right) = \frac{N_n(\bar{B})-n\mu(\bar{B})}{\mu(\complement\bar{B})}\,.\quad\square$$

Let us make at this place a remark concerning the covariance structure of $(N_n(B))_{B\in\mathcal{B}}$ supplementing the properties (2)-(4) on page 2:

It is easy to check that for any $B_i \in \mathcal{B}$, $i=1,2$,

(5) $\mathbb{E}(N_n(B_1)N_n(B_2)) = n\mu(B_1 \cap B_2) + n(n-1)\mu(B_1)\mu(B_2)$,

whence $\mathbb{E}(N_n(B_1)N_n(B_2)) = n(n-1)\mu(B_1)\mu(B_2)$ if $B_1 \cap B_2 = \emptyset$;

together with $\mathbb{E}(N_n(B_1))\mathbb{E}(N_n(B_2)) = n^2\mu(B_1)\mu(B_2)$ this yields

$\mathrm{cov}(N_n(B_1), N_n(B_2)) = -n\mu(B_1)\mu(B_2) \neq 0$ if $B_1 \cap B_2 = \emptyset$ and $\mu(B_i) > 0$, $i=1,2$;

therefore, $B_1 \cap B_2 = \emptyset$ does <u>not</u> imply that $N_n(B_1)$ and $N_n(B_2)$ are independent. (For the uniform empirical process, to be considered later, this implies that it is <u>not</u> a process with independent increments.) This situation changes if one considers instead the following

(6) <u>POISSONIZATION:</u> Let ν be a Poisson random variable (defined on the same p-space as the ξ_i's) with parameter λ and let for $B \in \mathcal{B}$

$$M(B) \equiv M(B,\omega) := \sum_{i=1}^{\nu(\omega)} 1_B(\xi_i(\omega)), \quad \omega \in \Omega.$$

Assume that ν is independent of the sequence $(\xi_i)_{i \in \mathbb{N}}$.

Then, for any pairwise disjoint $B_j \in \mathcal{B}$, $j=1,\dots,s$, the random variables $M(B_j)$, $j=1,\dots,s$, are independent.

Furthermore, for any $B \in \mathcal{B}$ and any $k \in \{0,1,2,\dots\}$ one has

$$\mathbb{P}(M(B)=k) = \frac{(\lambda\mu(B))^k}{k!} \exp(-\lambda\mu(B)).$$

<u>Proof.</u> Let us prove first the last statement:

$$\mathbb{P}(M(B)=k) = \mathbb{P}\Big(\bigcup_{\ell \geq k} \Big\{\sum_{i=1}^{\ell} 1_B(\xi_i)=k,\ \nu=\ell\Big\}\Big) = \sum_{\ell \geq k} \mathbb{P}\Big(\sum_{i=1}^{\ell} 1_B(\xi_i)=k\Big)\,\mathbb{P}(\nu=\ell)$$

$$= \sum_{\ell \geq k} \binom{\ell}{k} \mu(B)^k (1-\mu(B))^{\ell-k} \frac{\lambda^\ell}{\ell!} \exp(-\lambda)$$

$$= \sum_{\ell \geq k} \frac{\ell!}{k!(\ell-k)!} \mu(B)^k (1-\mu(B))^{\ell-k} \frac{\lambda^{\ell-k}\lambda^k}{\ell!} \exp(-\lambda)$$

$$\overset{(\ell-k=:m)}{=} \frac{(\lambda\mu(B))^k}{k!} \exp(-\lambda) \left[\sum_{m \geq 0} \frac{(\lambda(1-\mu(B)))^m}{m!}\right] = \frac{(\lambda\mu(B))^k}{k!} \exp(-\lambda\mu(B)).$$

As to the independence assertion let $B_{s+1} := \mathcal{C}(\bigcup_{j=1}^{s} B_j)$, $k := \sum_{j=1}^{s} k_j$, and $k_{s+1} := \ell - k$ for $\ell \geq k$. Then

$$\mathbb{P}(M(B_j) = k_j,\ j=1,\dots,s) = \mathbb{P}(\bigcup_{\ell\geq k}\{\sum_{i=1}^{\ell} 1_{B_j}(\xi_i) = k_j,\ j=1,\dots,s+1;\ \nu=\ell\})$$

$$= \sum_{\ell\geq k} \mathbb{P}(\sum_{i=1}^{\ell} 1_{B_j}(\xi_i) = k_j,\ j=1,\dots,s+1)\,\mathbb{P}(\nu=\ell)$$

$$= \sum_{\ell\geq k} \frac{\ell!}{k_1!\dots k_s!(\ell-k)!}\ \mu(B_1)^{k_1}\dots\mu(B_s)^{k_s}\mu(B_{s+1})^{\ell-k}\ \frac{\lambda^{k_1+\dots+k_s}\lambda^{\ell-k}}{\ell!}\ \exp(-\lambda)$$

$$= \prod_{j=1}^{s} \frac{(\lambda\mu(B_j))^{k_j}}{k_j!}\ \exp(-\lambda)\cdot\left[\sum_{m\geq 0}\frac{(\lambda\mu(B_{s+1}))^m}{m!}\right]\dots\dots(*),\ \text{where}$$

$[\dots.] = \exp(\lambda\mu(B_{s+1}))$, whence

$$\exp(-\lambda)\ [\dots.] = \exp(-\lambda(\sum_{j=1}^{s+1}\mu(B_j)))\ \exp(\lambda\mu(B_{s+1})) = \exp(-\lambda(\sum_{j=1}^{s}\mu(B_j))).$$

Therefore

$$(*) = \prod_{j=1}^{s} \frac{(\lambda\mu(B_j))^{k_j}}{k_j!}\ \exp(-\lambda\mu(B_j)). \qquad \square$$

Later we will consider for a given $\mathcal{C}\subset\mathcal{B}$ the so-called EMPIRICAL $\mathcal{C}$-PROCESS $\beta_n \equiv (\beta_n(C))_{C\in\mathcal{C}}$ defined by

$$\beta_n(C) := n^{1/2}(\mu_n(C)-\mu(C)),\ C\in\mathcal{C}.$$

Using (5) one obtains

$$\operatorname{cov}(\beta_n(C_1),\ \beta_n(C_2)) = \mu(C_1\cap C_2) - \mu(C_1)\mu(C_2),\ C_1,C_2\in\mathcal{C}.$$

Furthermore, $n^{1/2}(\beta_n(C_1) - \beta_n(C_2)) = \sum_{i=1}^{n} \eta_i(C_1,C_2)$ with

$\eta_i \equiv \eta_i(C_1,C_2) := 1_{C_1}(\xi_i) - 1_{C_2}(\xi_i) - (\mu(C_1) - \mu(C_2))$ being independent and identically distributed with $\mathbb{E}(\eta_i) = 0$ and

$\operatorname{Var}(\eta_i) = \mu(C_1\Delta C_2) - (\mu(C_1) - \mu(C_2))^2 \leq \mu(C_1\Delta C_2)$, whence the following Bernstein-type inequality applies (cf. G. Bennett (1962)):

(7) Let $\eta_1,\eta_2,\ldots$ be a sequence of independent random variables with $\mathbb{E}(\eta_i) = 0$ and $\mathrm{Var}(\eta_i) = \sigma_i^2$ and suppose that $\sup|\eta_i| \leqq M$ for some constant $0<M<\infty$; let $S_n := \sum_{i=1}^{n} \eta_i$ and $\tau_n^2 := \sum_{i=1}^{n} \sigma_i^2$; then for all n and $\varepsilon>0$

$$\mathbb{P}(S_n \geqq \varepsilon) \leqq \exp\left(-\frac{\varepsilon^2/2}{\tau_n^2+\varepsilon M/3}\right).$$

From (7) we obtain immediately

LEMMA 4. For every n and a>0 one has for any $C_i \in \mathcal{C}$, i=1,2,

(i) $$\mathbb{P}(|\beta_n(C_1)-\beta_n(C_2)|\geqq a) \leqq 2\exp\left(-\frac{na^2}{2n\mu(C_1\Delta C_2)+4n^{1/2}a/3}\right)$$

and for any $C\in\mathcal{C}$

(ii) $$\mathbb{P}(|\beta_n(C)|\geqq a) \leqq 2\exp\left(-\frac{a^2}{2\mu(C)(1-\mu(C))+an^{-1/2}}\right).$$

We will conclude this section with a further fundamental property concerning the so-called EMPIRICAL $\mathcal{C}$-DISCREPANCY

$$D_n(\mathcal{C},\mu) := \sup_{C\in\mathcal{C}} |\mu_n(C)-\mu(C)|.$$

In what follows we shall write $\|\mu_n-\mu\|$ instead of $D_n(\mathcal{C},\mu)$ and we assume that $\|\mu_n-\mu\|$ is a random variable, (i.e. $\mathcal{F},\mathcal{B}$-measurable). Then:

LEMMA 5. $(\|\mu_n-\mu\|)_{n\in\mathbb{N}}$ is a REVERSED SUBMARTINGALE w.r.t. the sequence of σ-fields $\mathcal{G}_n := \sigma(\{\mu_n(B), \mu_{n+1}(B),\ldots : B\in\mathcal{B}\})$ which means that for each m, n with $m\leqq n$

$$\mathbb{E}(\|\mu_m-\mu\| \,|\, \mathcal{G}_n) \geqq \|\mu_n-\mu\| \quad \mathbb{P}\text{-a.s.}$$

Proof. As shown in Gaenssler-Stute (1977), 6.5.5(c), the following holds: For each $C\in\mathcal{C}$ the process $(\mu_n(C)-\mu(C))_{n\in\mathbb{N}}$ is a REVERSED MARTINGALE w.r.t. $\mathcal{G}_n$, i.e., for each m, n with $m\leqq n$ one has

$$\mathbb{E}((\mu_m(C)-\mu(C)) \,|\, \mathcal{G}_n) = \mu_n(C)-\mu(C);$$

therefore

$$\mathbb{E}(\sup_{C\in\mathcal{C}} |\mu_m(C)-\mu(C)| \,|\, \mathcal{G}_n)$$

$$\geq \sup_{C\in\mathcal{C}} |\,\mathbb{E}((\mu_m(C)-\mu(C))|\mathcal{G}_n)\,| = \sup_{C\in\mathcal{C}} |\mu_n(C)-\mu(C)|. \quad \square$$

Now, as in the case of submartingales, there holds an analogous CONVERGENCE THEOREM FOR REVERSED SUBMARTINGALES (cf. Gaenssler-Stute (1977), 6.5.10) stating that for any reversed submartingale $(T_n)_{n\in\mathbb{N}}$ (on some p-space $(\Omega,\mathcal{F},\mathbb{P})$) w.r.t. a monotone decreasing sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of sub-σ-fields of $\mathcal{F}$ satisfying the condition that $\inf_n \mathbb{E}(T_n)>-\infty$ there exists an integrable random variable T_∞ such that $T_n\to T_\infty$ $\mathbb{P}$-a.s. and in the mean.

From this and Lemma 5 one obtains a rather simple proof of the following result (cf. D. Pollard (1981)) which, in a similar form, was one of the main results in Steele's paper (cf. M. Steele (1978)) proved there with different methods based on the ergodic theory of subadditive stochastic processes.

LEMMA 6. Let $(\nu_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of non-negative integer valued random variables on $(\Omega,\mathcal{F},\mathbb{P})$ such that $\nu_n \overset{\mathbb{P}}{\to} \infty$ (where $\overset{\mathbb{P}}{\to}$ denotes convergence in probability; also here and in the following all statements about convergence are understood to hold as n tends to infinity). Then

$$\|\mu_n-\mu\|\to 0 \;\; \mathbb{P}\text{-a.s. iff } \; \|\mu_{\nu_n}-\mu\| \overset{\mathbb{P}}{\to} 0;$$

in particular, $\|\mu_n-\mu\|\to 0$ $\mathbb{P}$-a.s. iff $\|\mu_n-\mu\| \overset{\mathbb{P}}{\to} 0.$

(Note that according to our measurability assumption on $\|\mu_n-\mu\|$ also the RANDOMIZED DISCREPANCY $\|\mu_{\nu_n}-\mu\|$ is a random variable; in fact,

$$\{\omega: \|\mu_{\nu_n(\omega)}(\cdot,\omega)-\mu\|\leq a\} = \bigcup_{j\in\mathbb{Z}_+} \{\nu_n=j\}\cap\{\|\mu_j-\mu\|\leq a\} \quad \text{for each } a\geq 0.)$$

Proof. 1.) Only if-part: $\nu_n \overset{\mathbb{P}}{\to} \infty$ implies that for any subsequence $(\nu_{n'})$ of (ν_n) there exists a further subsequence $(\nu_{n''})$ such that $\nu_{n''}\to\infty$ $\mathbb{P}$-a.s., whence $\|\mu_{\nu_{n''}}-\mu\|\to 0$ $\mathbb{P}$-a.s. as n" tends to infinity, and therefore $\|\mu_{\nu_n}-\mu\| \overset{\mathbb{P}}{\to} 0$.

2.) If-part: According to Lemma 5 the process $(\|\mu_n-\mu\|,\mathcal{G}_n)_{n\in\mathbb{N}}$ is a reversed submartingale. It is uniformly bounded; therefore, by the convergence theorem for reversed submartingales mentioned before, there exists an integrable random variable T_∞ such that $\|\mu_n - \mu\| \to T_\infty$ $\mathbb{P}$-a.s. From this it follows as in part 1.) of our proof that $\|\mu_{\nu_n} - \mu\| \overset{\mathbb{P}}{\to} T_\infty$, whence, by assumption, it follows that $T_\infty = 0$ $\mathbb{P}$-a.s. □