Reading group mathematical foundations of statistics

Chapter 3.6: Random Sample Size, Poissonization, and Kac Processes

Stefan Franssen, Msc

October 5, 2020

Set up

Suppose we have an infinite collection of random variables

$$X_1, X_2, \ldots \overset{\text{iid}}{\sim} P.$$

Now suppose we observe some random number of observations N_n, possibly dependent on the observations, will we still get the Donsker theorems?

Random Sample Size

Theorem

Let $\mathcal{F}$ be a Donsker class of measurable functions. Suppose that N_n is a sequence of positive, integer valued random variables such that $N_n/c_n \to \nu$ in probability, for a random variable ν with $P(\nu > 0) = 1$ and a deterministic sequence $c_n \to \infty$. Then the sequence $\mathbb{G}_{N_n}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight Brownian bridge as $n \to \infty$.

Proof part 1

First assume $c_n = n$ and $N_n/n \leq M$ for some number M.

Proof part 1

First assume $c_n = n$ and $N_n/n \leq M$ for some number M. Define

$$\mathbb{Z}_n(s, f) = n^{-1/2} \sum_{i=1}^{\lfloor ns \rfloor} (\delta_{X_i} - P)$$

for $s \in [0, M]$.

Proof part 1

First assume $c_n = n$ and $N_n/n \leq M$ for some number M. Define

$$\mathbb{Z}_n(s, f) = n^{-1/2} \sum_{i=1}^{\lfloor ns \rfloor} (\delta_{X_i} - P)$$

for $s \in [0, M]$. This is again a partial sum process, so we can apply the same tools as in 2.12, namely a small extension of theorem 2.12.1, to conclude that this converges to a Kiefer-Muller process $\mathbb{Z}$.

Proof part 2

Then by an upcoming lemma, when $c_n = n$ and $0 \leq \frac{N_n}{n} \leq M$, the sequence $(\mathbb{Z}_n, \frac{N_n}{n})$ converges weakly in $\ell([0, M] \times \mathcal{F}) \times \mathbb{R}$ to a pair $(\mathbb{Z}, \nu)$ of independent random elements $\mathbb{Z}$ and ν.

Proof part 2

Then by an upcoming lemma, when $c_n = n$ and $0 \leq \frac{N_n}{n} \leq M$, the sequence $(\mathbb{Z}_n, \frac{N_n}{n})$ converges weakly in $\ell([0, M] \times \mathcal{F}) \times \mathbb{R}$ to a pair $(\mathbb{Z}, \nu)$ of independent random elements $\mathbb{Z}$ and ν. So conclude

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N_n} (\delta_{X_i} - P) f = \mathbb{Z}\left(\frac{N_n}{n}, f\right) =: g\left(\mathbb{Z}_n, \frac{N_n}{n}\right) f.$$

Proof part 2

Then by an upcoming lemma, when $c_n = n$ and $0 \leq \frac{N_n}{n} \leq M$, the sequence $(\mathbb{Z}_n, \frac{N_n}{n})$ converges weakly in $\ell([0, M] \times \mathcal{F}) \times \mathbb{R}$ to a pair $(\mathbb{Z}, \nu)$ of independent random elements $\mathbb{Z}$ and ν. So conclude

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N_n} (\delta_{X_i} - P) f = \mathbb{Z}\left(\frac{N_n}{n}, f\right) =: g\left(\mathbb{Z}_n, \frac{N_n}{n}\right) f.$$

Then the map g is continuous almost surely. So we can apply the continuous mapping theorem, and conclude that

$$G_{N_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{N_n} (\delta_{X_i} - P) \rightsquigarrow \frac{1}{\sqrt{\nu}} \mathbb{Z}(\nu, \cdot).$$

Proof part 2

Then by an upcoming lemma, when $c_n = n$ and $0 \leq \frac{N_n}{n} \leq M$, the sequence $(\mathbb{Z}_n, \frac{N_n}{n})$ converges weakly in $\ell([0, M] \times \mathcal{F}) \times \mathbb{R}$ to a pair $(\mathbb{Z}, \nu)$ of independent random elements $\mathbb{Z}$ and ν. So conclude

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N_n} (\delta_{X_i} - P)f = \mathbb{Z}\left(\frac{N_n}{n}, f\right) =: g\left(\mathbb{Z}_n, \frac{N_n}{n}\right) f.$$

Then the map g is continuous almost surely. So we can apply the continuous mapping theorem, and conclude that

$$G_{N_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{N_n} (\delta_{X_i} - P) \rightsquigarrow \frac{1}{\sqrt{\nu}} \mathbb{Z}(\nu, \dot{)}.$$

Since ν and $\mathbb{Z}$ are independent, and $\nu^{-1/2}\mathbb{Z}(\nu, \cdot)$ is distributed as a Brownian bridge for every deterministic ν, the variable on the right hand side of the display is distributed as a Brownian bridge.

proof part 3

If $\frac{N_n}{n}$ is not bounded, define $M_{n,M} = N_n \wedge (Mn)$. Then $N_{n,M}$ is bounded, so we can apply the previous argument to $\mathbb{G}_{N_{n,M}}$ to conclude that $\mathbb{G}_{N_{n,M}} \rightsquigarrow \mathbb{G}$.

proof part 3

If $\frac{N_n}{n}$ is not bounded, define $M_{n,M} = N_n \wedge (Mn)$. Then $N_{n,M}$ is bounded, so we can apply the previous argument to $\mathbb{G}_{N_{n,M}}$ to conclude that $\mathbb{G}_{N_{n,M}} \rightsquigarrow \mathbb{G}$. The probability that $N_{n,M}$ is different from N_n can be made arbitrarily small by increasing M, thus the result also follows in case of unbounded N_n/n.

proof part 3

If $\frac{N_n}{n}$ is not bounded, define $M_{n,M} = N_n \wedge (Mn)$. Then $N_{n,M}$ is bounded, so we can apply the previous argument to $\mathbb{G}_{N_{n,M}}$ to conclude that $\mathbb{G}_{N_{n,M}} \rightsquigarrow \mathbb{G}$. The probability that $N_{n,M}$ is different from N_n can be made arbitrarily small by increasing M, thus the result also follows in case of unbounded N_n/n. The case where c_n is not equal to n can be done by relabeling. If c_n is not an integer, we can study $\lfloor c_n \rfloor$ instead. Therefore, assume each c_n is an integer.

proof part 3

If $\frac{N_n}{n}$ is not bounded, define $M_{n,M} = N_n \wedge (Mn)$. Then $N_{n,M}$ is bounded, so we can apply the previous argument to $\mathbb{G}_{N_{n,M}}$ to conclude that $\mathbb{G}_{N_{n,M}} \rightsquigarrow \mathbb{G}$. The probability that $N_{n,M}$ is different from N_n can be made arbitrarily small by increasing M, thus the result also follows in case of unbounded N_n/n. The case where c_n is not equal to n can be done by relabeling. If c_n is not an integer, we can study $\lfloor c_n \rfloor$ instead. Therefore, assume each c_n is an integer. For every subsequence of (c_n), we find a further subsequence c_n' which is strictly increasing. So without loss of generality assume c_n is strictly increasing.

proof part 3

If $\frac{N_n}{n}$ is not bounded, define $M_{n,M} = N_n \wedge (Mn)$. Then $N_{n,M}$ is bounded, so we can apply the previous argument to $\mathbb{G}_{N_{n,M}}$ to conclude that $\mathbb{G}_{N_{n,M}} \rightsquigarrow \mathbb{G}$. The probability that $N_{n,M}$ is different from N_n can be made arbitrarily small by increasing M, thus the result also follows in case of unbounded N_n/n. The case where c_n is not equal to n can be done by relabeling. If c_n is not an integer, we can study $\lfloor c_n \rfloor$ instead. Therefore, assume each c_n is an integer. For every subsequence of (c_n), we find a further subsequence c_n' which is strictly increasing. So without loss of generality assume c_n is strictly increasing. Then define $N_k' = N_n$ if $c_n = k$ and $N_k' = k\nu$ if $k \neq c_n$ for every n. Then $N_k'/k \xrightarrow{P} \nu$, hence $G_{N_k'} \rightsquigarrow G$. The sequence $\mathbb{G}_{N_n}$ is a subsequence.

The Lemma

Lemma

Let $\mathcal{F}$ be a Donsker class and ν_n a sequence of random variables such that $\nu_n \xrightarrow{P} \nu$ for a random variable ν. Then the sequence of sequential empirical processes $\mathbb{Z}_n$ satisfies $(\mathbb{Z}_n, \nu_n) \rightsquigarrow (\mathbb{Z}, \nu)$ in $\ell^\infty([0, M] \times \mathcal{F}) \times \mathbb{R}$, where $\mathbb{Z}$ and ν are independent.

proof Part 1

Let $k_n \to \infty$ slowly enough that $k_n = o(\sqrt{n})$. Set

$$\mathbb{Z}'_n(s, f) = \frac{1}{\sqrt{n}} \sum_{i=k_n+1}^{\lfloor ns \rfloor} (\delta_{X_i} - P).$$

proof Part 1

Let $k_n \to \infty$ slowly enough that $k_n = o(\sqrt{n})$. Set

$$\mathbb{Z}'_n(s, f) = \frac{1}{\sqrt{n}} \sum_{i=k_n+1}^{\lfloor ns \rfloor} (\delta_{X_i} - P).$$

Then $\mathbb{Z}_n - \mathbb{Z}'_n$ converges, in $\ell^\infty([0, M] \times \mathcal{F})$, to zero by Slutsky's Lemma.

proof Part 1

Let $k_n \to \infty$ slowly enough that $k_n = o(\sqrt{n})$. Set

$$\mathbb{Z}'_n(s,f) = \frac{1}{\sqrt{n}} \sum_{i=k_n+1}^{\lfloor ns \rfloor} (\delta_{X_i} - P).$$

Then $\mathbb{Z}_n - \mathbb{Z}'_n$ converges, in $\ell^\infty([0,M] \times \mathcal{F})$, to zero by Slutsky's Lemma. Thus again by Slutsky's Lemma, $\mathbb{Z}'_n$ has the same limit as $\mathbb{Z}_n$.

proof part 2

By Doobs martingale convergence theorem,

$$\mathbb{P}(\nu \in B | X_1, \dots, X_k) \to \mathbb{P}(\nu \in B | X_1, X_2, \dots)$$

in mean as $k \to \infty$.

proof part 2

By Doobs martingale convergence theorem,

$$\mathbb{P}(\nu \in B | X_1, \dots, X_k) \to \mathbb{P}(\nu \in B | X_1, X_2, \dots)$$

in mean as $k \to \infty$. Therefore

$$\lim_{n\to\infty} \mathbb{P}^* \left(\mathbb{Z}'_n \in A, \nu \in B\right) = \lim_{n\to\infty} \mathbb{E}\left(\mathbb{1}_A(\mathbb{Z}'_n)^* \mathbb{P}(\nu \in B | X_1, \dots, X_{k_n})\right).$$

Since $\mathbb{Z}'_n$ is independent of $X_1, \dots, X_{K_n}$, we can factorize this expectation as

$$\mathbb{P}(\mathbb{Z}'_n \in A)\mathbb{P}(\nu \in B).$$

This converges to $\mathbb{P}(\mathbb{Z} \in A)\mathbb{P}(\nu \in B)$ for every continuity set A, which concludes the proof.

Equivalence of limits in probability

Lemma

Let $\mathcal{F}$ be a Donsker class of measurable functions. Suppose that N_n is a sequence of positive, integer-valued random variables such that $N_n/n \xrightarrow{P} 1$. Then the sequence $G_{N_n} - G_n$ converges in outer probability to zero in $\ell^\infty(\mathcal{F})$ as $n \to \infty$.

proof

With the same notation in the previous proofs, the sequence $(\mathbb{Z}_n, \frac{N_n}{n})$ converges in distribution to $(\mathbb{Z}, 1)$. By the continuous mapping hteorem, $\mathbb{Z}_n(\frac{N_n}{n}, \cdot) - \mathbb{Z}_n(1, \cdot) \rightsquigarrow \mathbb{Z}(1, \cdot) - \mathbb{Z}(1, \cdot) = 0$. Convergence in distribution and in outer probability to degenerate limit are equivalent.

Break

The Kac Processes

Let the sample size N_n be a Poisson random variable with mean n, independent of the i.i.d. observations $X_1, X_2, \ldots$.

The Kac Processes

Let the sample size N_n be a Poisson random variable with mean n, independent of the i.i.d. observations $X_1, X_2, \ldots$. The Kac empirical point process is the random measure given by

$$\mathbb{N} = \sum_{i=1}^{N_n} \delta_{X_i}.$$

The Kac Processes

Let the sample size N_n be a Poisson random variable with mean n, independent of the i.i.d. observations $X_1, X_2, \ldots$. The Kac empirical point process is the random measure given by

$$\mathbb{N} = \sum_{i=1}^{N_n} \delta_{X_i}.$$

For every measurable set C, the random variable $\mathbb{N}_n(C)$ is poisson distributed with mean $nP(c)$.

The Kac Processes

Let the sample size N_n be a Poisson random variable with mean n, independent of the i.i.d. observations $X_1, X_2, \dots$. The Kac empirical point process is the random measure given by

$$\mathbb{N} = \sum_{i=1}^{N_n} \delta_{X_i}.$$

For every measurable set C, the random variable $\mathbb{N}_n(C)$ is poisson distributed with mean $nP(c)$. For disjoint $C_1, \dots, C_k$, the random variables $\mathbb{N}_n(C_1), \dots, \mathbb{N}_n(C_k)$ are indepedent.

Standardizing

For a class of measurable functions $\mathcal{F}$, consider

$$\{\mathbb{N}_n(f) : f \in \mathcal{F}\}.$$

Standardizing

For a class of measurable functions $\mathcal{F}$, consider

$$\{\mathbb{N}_n(f) : f \in \mathcal{F}\}.$$

The mean and variance of this proces satisfy

$$\mathbb{E}[\mathbb{N}_n(f)] = nPf = \mathrm{Var}(\mathbb{N}_n(f)).$$

Standardizing

For a class of measurable functions $\mathcal{F}$, consider

$$\{\mathbb{N}_n(f) : f \in \mathcal{F}\}.$$

The mean and variance of this proces satisfy

$$\mathbb{E}[\mathbb{N}_n(f)] = nPf = \mathrm{Var}(\mathbb{N}_n(f)).$$

So we can standardize in the following way

$$\mathbb{Z}_n = \frac{1}{\sqrt{n}}\left(\mathbb{N}_n - nP\right) = \sqrt{\frac{N_n}{n}}\mathbb{G}_{N_n} + \sqrt{n}\left(\frac{N_n}{n} - 1\right)P.$$

Standardizing

For a class of measurable functions $\mathcal{F}$, consider

$$\{\mathbb{N}_n(f) : f \in \mathcal{F}\}.$$

The mean and variance of this proces satisfy

$$\mathbb{E}[\mathbb{N}_n(f)] = nPf = \mathrm{Var}(\mathbb{N}_n(f)).$$

So we can standardize in the following way

$$\mathbb{Z}_n = \frac{1}{\sqrt{n}}(\mathbb{N}_n - nP) = \sqrt{\frac{N_n}{n}}\mathbb{G}_{N_n} + \sqrt{n}\left(\frac{N_n}{n} - 1\right)P.$$

A collection of functions $\mathcal{F}$ is called P-Kac if the sequence $\mathbb{Z}_n$ converges in distribution to a tight limit process in $\ell^\infty(\mathcal{F})$.

Donsker + $\|P\|_{\mathcal{F}} < \infty$ implies Kac

Since N_n converges weakly and thus in probability to 1, we can apply the last theoremm which implies

$$\mathbb{G}_{N_n} - \mathbb{G}_n \xrightarrow{P} 0,$$

if $\mathcal{F}$ is Donsker.

Donsker + $\|P\|_{\mathcal{F}} < \infty$ implies Kac

Since N_n converges weakly and thus in probability to 1, we can apply the last theoremm which implies

$$\mathbb{G}_{N_n} - \mathbb{G}_n \xrightarrow{P} 0,$$

if $\mathcal{F}$ is Donsker. By Slutsky's Lemma, this means that the standardized process converges in case $\sqrt{n}\left(\frac{N_n}{n} - 1\right) P$ converges.

Donsker + $\|P\|_{\mathcal{F}} < \infty$ implies Kac

Since N_n converges weakly and thus in probability to 1, we can apply the last theoremm which implies

$$\mathbb{G}_{N_n} - \mathbb{G}_n \xrightarrow{P} 0,$$

if $\mathcal{F}$ is Donsker. By Slutsky's Lemma, this means that the standardized process converges in case $\sqrt{n}\left(\frac{N_n}{n} - 1\right) P$ converges. The latter limit exists as soon as $\|P\|_{\mathcal{F}}$ is finite.

Donsker + $\|P\|_{\mathcal{F}} < \infty$ implies Kac

Since N_n converges weakly and thus in probability to 1, we can apply the last theoremm which implies

$$\mathbb{G}_{N_n} - \mathbb{G}_n \xrightarrow{P} 0,$$

if $\mathcal{F}$ is Donsker. By Slutsky's Lemma, this means that the standardized process converges in case $\sqrt{n}\left(\frac{N_n}{n} - 1\right) P$ converges. The latter limit exists as soon as $\|P\|_{\mathcal{F}}$ is finite. The limit of this process is given by

$$\mathbb{G} + ZP$$

For a Brownian bridge $\mathbb{G}$ and an independent standard gaussian random variable Z.

Equivalent processes

Since $\mathbb{N}_n$ is a Poisson process with intensity measure nP, it can be written as the sum of n i.i.d. Poisson processes of intensity measure P. Let $Y_1, Y_2, \ldots$ be an $i.i.d.$ sequence of Poisson(1) variables, and let $X_{i,j}$ be an array of i.i.d. copies of X_1.

Equivalent processes

Since $\mathbb{N}_n$ is a Poisson process with intensity measure nP, it can be written as the sum of n i.i.d. Poisson processes of intensity measure P. Let $Y_1, Y_2, \ldots$ be an $i.i.d.$ sequence of Poisson(1) variables, and let $X_{i,j}$ be an array of i.i.d. copies of X_1. Then the process

$$H_n = \sum_{i=1}^{n} (\delta_{X_{i,j}} - P)$$

is equal in distribution to $H_n' = \sum_{i=1}^{N_n} (\delta_{X_i} - P)$.

Equivalent processes

Since $\mathbb{N}_n$ is a Poisson process with intensity measure nP, it can be written as the sum of n i.i.d. Poisson processes of intensity measure P. Let $Y_1, Y_2, \ldots$ be an $i.i.d.$ sequence of Poisson(1) variables, and let $X_{i,j}$ be an array of i.i.d. copies of X_1. Then the process

$$H_n = \sum_{i=1}^{n} (\delta_{X_{i,j}} - P)$$

is equal in distribution to $H_n' = \sum_{i=1}^{N_n} (\delta_{X_i} - P)$. It follows therefore that the random-sample central limit theorem for $\mathbb{G}_{N_n}$ is equivalent to the central limit theorem for a deterministic number of Poisson processes of the type $\sum_{i=1}^{Y_i} (\delta_{X_{i,j}} - P)$.

Equivalent processes

Since $\mathbb{N}_n$ is a Poisson process with intensity measure nP, it can be written as the sum of n i.i.d. Poisson processes of intensity measure P. Let $Y_1, Y_2, \ldots$ be an $i.i.d.$ sequence of Poisson(1) variables, and let $X_{i,j}$ be an array of i.i.d. copies of X_1. Then the process

$$H_n = \sum_{i=1}^{n} (\delta_{X_{i,j}} - P)$$

is equal in distribution to $H'_n = \sum_{i=1}^{N_n} (\delta_{X_i} - P)$. It follows therefore that the random-sample central limit theorem for $\mathbb{G}_{N_n}$ is equivalent to the central limit theorem for a deterministic number of Poisson processes of the type $\sum_{i=1}^{Y_i} (\delta_{X_{i,j}} - P)$. Le Cam's Lemma compares the concentration of these processes with the concentration of the empirical process.

le Cam's Lemma

Lemma

Let N_n be poisson distributed with mean n and independent of the i.i.d. stochastic processes $Z_1, Z_2, \dots$. Then for any class of functions $\mathcal{F}$:

$$(1 - \frac{1}{e})\mathbb{E}^* \| \sum_{i=1}^{n} Z_i \|_{\mathcal{F}} \leq \mathbb{E}^* \| \sum_{i=1}^{N_n} Z_i \|_{\mathcal{F}}$$

Proof

The expected value of $\max(Y_i, 1)$, where Y_i is a Poisson(1) random variable, is $1 - \frac{1}{e}$.

Proof

The expected value of $\max(Y_i, 1)$, where Y_i is a Poisson(1) random variable, is $1 - \frac{1}{e}$.

$$(1 - \frac{1}{e})\mathbb{E}^* \| \sum_{i=1}^{n} Z_i \|$$

Proof

The expected value of $\max(Y_i, 1)$, where Y_i is a Poisson(1) random variable, is $1 - \frac{1}{e}$.

$$(1 - \frac{1}{e})\mathbb{E}^* \| \sum_{i=1}^{n} Z_i \| = \mathbb{E}_Z^* \| \mathbb{E}_Y \sum_{i=1}^{n} \max(Y_i, 1) Z_i \|_{\mathcal{F}}$$

Proof

The expected value of $\max(Y_i, 1)$, where Y_i is a Poisson(1) random variable, is $1 - \frac{1}{e}$.

$$(1 - \frac{1}{e})\mathbb{E}^* \| \sum_{i=1}^{n} Z_i \| = \mathbb{E}_Z^* \| \mathbb{E}_Y \sum_{i=1}^{n} \max(Y_i, 1) Z_i \|_{\mathcal{F}}$$

$$\leq \mathbb{E}^* \| \sum_{i=1}^{n} \max(Y_i, 1) Z_i \|_{\mathcal{F}}$$

Proof

The expected value of $\max(Y_i, 1)$, where Y_i is a Poisson(1) random variable, is $1 - \frac{1}{e}$.

$$
\begin{aligned}
(1 - \frac{1}{e})\mathbb{E}^*\|\sum_{i=1}^{n} Z_i\| &= \mathbb{E}_Z^*\|\mathbb{E}_Y \sum_{i=1}^{n} \max(Y_i, 1) Z_i\|_{\mathcal{F}} \\
&\leq \mathbb{E}^*\|\sum_{i=1}^{n} \max(Y_i, 1) Z_i\|_{\mathcal{F}} \\
&\leq \mathbb{E}_Y \mathbb{E}_Z^*\|\sum_{i=1}^{n} \sum_{j=1}^{Y_i} Z_{i,j}\|_{\mathcal{F}}
\end{aligned}
$$

Proof

The expected value of $\max(Y_i, 1)$, where Y_i is a Poisson(1) random variable, is $1 - \frac{1}{e}$.

$$
\begin{aligned}
(1 - \frac{1}{e})\mathbb{E}^*\|\sum_{i=1}^{n} Z_i\| &= \mathbb{E}_Z^*\|\mathbb{E}_Y \sum_{i=1}^{n} \max(Y_i, 1) Z_i\|_{\mathcal{F}} \\
&\leq \mathbb{E}^*\|\sum_{i=1}^{n} \max(Y_i, 1) Z_i\|_{\mathcal{F}} \\
&\leq \mathbb{E}_Y \mathbb{E}_Z^*\|\sum_{i=1}^{n} \sum_{j=1}^{Y_i} Z_{i,j}\|_{\mathcal{F}} \\
&= \mathbb{E}^*\|\sum_{i=1}^{N_n} Z_i\|_{\mathcal{F}}
\end{aligned}
$$

Equivalence Kac and Donsker

Theorem

A class $\mathcal{F}$ of measurable functions with $\|P\|_{\mathcal{F}} < \infty$ is Kac if and only if it is Donsker. In that case,

$$\|\mathbb{G}_{N_n} - \mathbb{G}_n\|^*_{\mathcal{F}} = O_P(n^{-1/4})$$

proof part 1

Define $\mathcal{G}'_n = n^{-1/2} \sum_{i=1}^{N_n} (\delta_{X_i} - P)$.

proof part 1

Define $\mathcal{G}'_n = n^{-1/2} \sum_{i=1}^{N_n} (\delta_{X_i} - P)$. Define $k = N_n - n$. Then the difference between G'_n and G_n will be k terms $n^{-1/2}(\delta_{X_i} - P)$.

proof part 1

Define $\mathcal{G}'_n = n^{-1/2} \sum_{i=1}^{N_n} (\delta_{X_i} - P)$. Define $k = N_n - n$. Then the difference between G'_n and G_n will be k terms $n^{-1/2}(\delta_{X_i} - P)$. By chebyshev's inequality, we have $\mathbb{P}(|N_n - n| \geq M\sqrt{n}) \leq M^{-2}$.

proof part 1

Define $\mathcal{G}'_n = n^{-1/2} \sum_{i=1}^{N_n} (\delta_{X_i} - P)$. Define $k = N_n - n$. Then the difference between G'_n and G_n will be k terms $n^{-1/2}(\delta_{X_i} - P)$. By chebyshev's inequality, we have $\mathbb{P}(|N_n - n| \geq M\sqrt{n}) \leq M^{-2}$. Therefore, for every $\epsilon > 0$ we get

$$\mathbb{P}^*(\|\mathbb{G}_n - \mathbb{G}'_n|_{\mathcal{F}} > \epsilon)$$

proof part 1

Define $\mathcal{G}'_n = n^{-1/2}\sum_{i=1}^{N_n}(\delta_{X_i} - P)$. Define $k = N_n - n$. Then the difference between G'_n and G_n will be k terms $n^{-1/2}(\delta_{X_i} - P)$. By chebyshev's inequality, we have $\mathbb{P}(|N_n - n| \geq M\sqrt{n}) \leq M^{-2}$. Therefore, for every $\epsilon > 0$ we get

$$\mathbb{P}^*(\|\mathbb{G}_n - \mathbb{G}'_n|_{\mathcal{F}} > \epsilon)$$

$$\leq \frac{1}{M^2} + \frac{1}{\epsilon}\sum_{|k|\leq M\sqrt{n}} \mathbb{P}(N_n = n + k)\mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^{|k|}(\delta_{X_i} - P)\|_{\mathcal{F}}$$

proof part 1

Define $\mathcal{G}'_n = n^{-1/2}\sum_{i=1}^{N_n}(\delta_{X_i} - P)$. Define $k = N_n - n$. Then the difference between G'_n and G_n will be k terms $n^{-1/2}(\delta_{X_i} - P)$. By chebyshev's inequality, we have $\mathbb{P}(|N_n - n| \geq M\sqrt{n}) \leq M^{-2}$. Therefore, for every $\epsilon > 0$ we get

$$
\begin{aligned}
&\mathbb{P}^*(\|\mathbb{G}_n - \mathbb{G}'_n|_{\mathcal{F}} > \epsilon) \\
&\leq \frac{1}{M^2} + \frac{1}{\epsilon}\sum_{|k|\leq M\sqrt{n}} \mathbb{P}(N_n = n + k)\mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^{|k|}(\delta_{X_i} - P)\|_{\mathcal{F}} \\
&\leq \frac{1}{M^2} + \frac{1}{\epsilon\sqrt{n}}\mathbb{E}^*\|\sum_{i=1}^{\lfloor M\sqrt{n}\rfloor}(\delta_{X_i} - P)\|_{\mathcal{F}}.
\end{aligned}
$$

proof part 2

If $\mathcal{F}$ is Donsker, then the sequence $n^{-1/4}\mathbb{E}^*\|\sum_{i=1}^{M\sqrt{n}}(\delta_{X_i} - P)\|_{\mathcal{F}}$ is bounded by Lemma 2.3.11.

proof part 2

If $\mathcal{F}$ is Donsker, then the sequence $n^{-1/4}\mathbb{E}^*\|\sum_{i=1}^{M\sqrt{n}}(\delta_{X_i} - P)\|_{\mathcal{F}}$ is bounded by Lemma 2.3.11. Then by the previous slide, $\mathbb{P}*(\|\mathbb{G}_n - \mathbb{G}'_n\|_{\mathcal{F}} > Kn^{-1/4})$ is bounded by $M^{-2} + K^{-1}O(1)$.

proof part 2

If $\mathcal{F}$ is Donsker, then the sequence $n^{-1/4}\mathbb{E}^*\|\sum_{i=1}^{M\sqrt{n}}(\delta_{X_i} - P)\|_{\mathcal{F}}$ is bounded by Lemma 2.3.11. Then by the previous slide, $\mathbb{P}*(\|\mathbb{G}_n - \mathbb{G}'_n\|_{\mathcal{F}} > Kn^{-1/4})$ is bounded by $M^{-2} + K^{-1}O(1)$. We can bound its limsup by choosing M and K large enough.

proof part 2

If $\mathcal{F}$ is Donsker, then the sequence $n^{-1/4}\mathbb{E}^*\|\sum_{i=1}^{M\sqrt{n}}(\delta_{X_i} - P)\|_{\mathcal{F}}$ is bounded by Lemma 2.3.11. Then by the previous slide, $\mathbb{P}*(\|\mathbb{G}_n - \mathbb{G}'_n\|_{\mathcal{F}} > Kn^{-1/4})$ is bounded by $M^{-2} + K^{-1}O(1)$. We can bound its limsup by choosing M and K large enough. Therefore $\|\mathbb{G}_n - \mathbb{G}'_n\|^*_{\mathcal{F}} = O_P(n^{-1/4})$.

proof part 2

If $\mathcal{F}$ is Donsker, then the sequence $n^{-1/4}\mathbb{E}^*\|\sum_{i=1}^{M\sqrt{n}}(\delta_{X_i} - P)\|_{\mathcal{F}}$ is bounded by Lemma 2.3.11. Then by the previous slide, $\mathbb{P}*(\|\mathbb{G}_n - \mathbb{G}'_n\|_{\mathcal{F}} > Kn^{-1/4})$ is bounded by $M^{-2} + K^{-1}O(1)$. We can bound its limsup by choosing M and K large enough. Therefore $\|\mathbb{G}_n - \mathbb{G}'_n\|^*_{\mathcal{F}} = O_P(n^{-1/4})$. Since $\|\mathbb{G}'_n - \mathbb{G}_{N_n}\| = (1 - \sqrt{\frac{N_n}{n}})\|\mathbb{G}'_n\|_{\mathcal{F}}$, we see that this is $O_P(n^{-1/2})$, and therefore the first direction follows.

proof part 3

If $\mathcal{F}$ is Kac, then $\mathbb{G}'_n = n^{-1/2}(\mathbb{N}_n - N_n P)$ converges in distribution to a tight limit process.

proof part 3

If $\mathcal{F}$ is Kac, then $\mathbb{G}'_n = n^{-1/2}(\mathbb{N}_n - N_n P)$ converges in distribution to a tight limit process. This means the standardized processes $Z_i = \sum_{j=1}^{Y_i}(\delta_{X_{i,j}} - P)$ satisfying the central limit theorem.

proof part 3

If $\mathcal{F}$ is Kac, then $\mathbb{G}'_n = n^{-1/2}(\mathbb{N}_n - N_n P)$ converges in distribution to a tight limit process. This means the standardized processes $Z_i = \sum_{j=1}^{Y_i}(\delta_{X_{i,j}} - P)$ satisfying the central limit theorem. Thus $n^{-1/2}\mathbb{E}^*\| \sum_{i=1}^{n} Z_i\|$ is bounded by Lemma 2.3.11.

proof part 3

If $\mathcal{F}$ is Kac, then $\mathbb{G}'_n = n^{-1/2}(\mathbb{N}_n - N_n P)$ converges in distribution to a tight limit process. This means the standardized processes $Z_i = \sum_{j=1}^{Y_i}(\delta_{X_{i,j}} - P)$ satisfying the central limit theorem. Thus $n^{-1/2}\mathbb{E}^*\|\sum_{i=1}^n Z_i\|$ is bounded by Lemma 2.3.11. By Le Cams Lemma these expectations give an upper bound, up to a positive constant factor, of $n^{-1/2}\mathbb{E}^*\|\sum_{i=1}^n(\delta_{x_i} - P)\|$.

proof part 3

If $\mathcal{F}$ is Kac, then $\mathbb{G}'_n = n^{-1/2}(\mathbb{N}_n - N_n P)$ converges in distribution to a tight limit process. This means the standardized processes $Z_i = \sum_{j=1}^{Y_i}(\delta_{X_{i,j}} - P)$ satisfying the central limit theorem. Thus $n^{-1/2}\mathbb{E}^*\|\sum_{i=1}^{n} Z_i\|$ is bounded by Lemma 2.3.11. By Le Cams Lemma these expectations give an upper bound, up to a positive constant factor, of $n^{-1/2}\mathbb{E}^*\|\sum_{i=1}^{n}(\delta_{x_i} - P)\|$. Then finish the proof as in the reverse direction and we are done.

The end