Elect. Comm. in Probab. **14** (2009), 464–473

ELECTRONIC COMMUNICATIONS in PROBABILITY

AN EXTENSION OF MINEKA'S COUPLING INEQUALITY

ANNA PÓSFAI
Analysis and Stochastics Research Group of the Hungarian Academy of Sciences, Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, Szeged 6720, Hungary
email: posfai@math.u-szeged.hu

Submitted May 18, 2009, *accepted in final form* October 8, 2009

AMS 2000 Subject classification: 60G50, 62E17
Keywords: Mineka coupling, translated compound Poisson approximation, total variation distance, coupon collecting

Abstract

In this paper we propose a refinement of Mineka's coupling inequality that gives a better upper bound for $d_{TV}(\mathscr{L}(W), \mathscr{L}(W+1))$, where W is a sum of n independent integer valued random variables, in the case when $\mathbf{Var}W \gg n$.

Introduction

Translated compound Poisson approximation of sums of independent integer valued random variables has been studied in a series of papers. Using Stein's method, [1] and [2] give bounds for the errors of such approximations in total variation distance, which is defined by

$$d_{TV}(\mathscr{L}(X), \mathscr{L}(Y)) := \sup_{A \subset Z} |\mathbf{P}(X \in A) - \mathbf{P}(Y \in A)|, \tag{1}$$

$Z := \{\dots, -1, 0, 1, \dots\}$. Their upper bounds are expressed with the help of the first three moments of the summands $X_1, X_2, \dots, X_n$ and the critical ingredient $d_{TV}\left(\mathscr{L}\left(W_n\right), \mathscr{L}\left(W_n+1\right)\right)$, where $W_n = \sum_{j=1}^n X_j$.

The expression $d_{TV}\left(\mathscr{L}\left(W_n\right), \mathscr{L}\left(W_n+1\right)\right)$ is usually bounded by the Mineka coupling [3] (p. 44), which typically yields a bound of order $1/\sqrt{n}$. If the X_j's are roughly similar in magnitude, this is comparable with the order $O(1/\sqrt{\mathbf{Var}W_n})$ expected for the error in the central limit theorem. However, if the distributions of the X_j become progressively more spread out as j increases, then $\mathbf{Var}W_n$ may grow faster than n, and then $1/\sqrt{n}$ is bigger than the ideal order $O(1/\sqrt{\mathbf{Var}W_n})$. In this paper we introduce a new coupling which allows us to improve the bounds obtained by the Mineka coupling in such cases.

An example for this situation is given by the coupon collector's problem: a collector samples with replacement a set of n coupons until the random time $W_{n,m}$ at which he obtains $n-m$ of them for the first time, $m \in \{0, 1, \dots, n-1\}$. Relying on the new technique, we shall show that

$d_{TV}\left(\mathscr{L}\left(W_{n,m}\right),\mathscr{L}\left(W_{n,m}+1\right)\right)=O\left(1/\sqrt{\mathbf{Var}W_{n,m}}\right)$, and therefore that a translated compound Poisson approximation to the collector's waiting time $W_{n,m}$, with ideal error rate, can be obtained in all ranges of n and m in which a central limit theorem can be proved.

The coupling and the bound for $d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1))$

As mentioned in the introduction, it may happen that $d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1)) = O\left(\frac{1}{\sqrt{\mathbf{Var}W}}\right)$, where $1/\sqrt{\mathbf{Var}W} \ll 1/\sqrt{n}$, and thus the usual coupling proof of such a relation does not work. In the following lemma we show that this result can be established for sums of iid discrete uniform random variables with the help of a new coupling.

Lemma. Let $U_1, U_2, \dots, U_r$, $r \geq 2$, be independent identically distributed random variables with discrete uniform distribution on $\{1, 2, \dots, 2l-1, 2l\}$ for some $l \geq 1$ integer. If $V_r = \sum_{j=1}^{r} U_j$, then

$$d_{TV}(\mathscr{L}(V_r),\mathscr{L}(V_r+1)) \leq \frac{1}{l\sqrt{r}}.$$

Proof. We construct a coupling of (V_r, V_r+1). Let U_1 be an arbitrary random variable of uniform distribution on $\{1, 2, \dots, 2l\}$. If $U_1 \in \{1, 2, \dots, 2l-1\}$, then define

$$U_1' = U_1 + 1 \quad \text{and} \quad U_j' = U_j,\ 2 \leq j \leq r,$$

where $U_1, \dots, U_r$ are independent; while if $U_1 = 2l$, then put

$$U_1' = 1 \quad \text{and} \quad U_j = \widetilde{U}_j + lI_j, \quad U_j' = \widetilde{U}_j + l(1-I_j), \quad 2 \leq j \leq r,$$

where $\widetilde{U}_j$ has uniform distribution on $\{1, \dots, l\}$, I_j takes on the values 0 and 1, each with probability $1/2$, and $\widetilde{U}_j$, I_j, $2 \leq j \leq r$, are independent, also of U_1.
Introducing $V_s := \sum_{j=1}^{s} U_j$ and $V_s' := \sum_{j=1}^{s} U_j'$, $s \in \{1, \dots, r\}$, we see that

$$S_s := (V_s+1) - V_s' = \begin{cases} 0, & \text{if } U_1 \in \{1, 2, \dots, 2l-1\} \\ 2l + \sum_{j=2}^{s}(U_j - U_j'), & \text{if } U_1 = 2l, \end{cases}$$

where
$$U_j - U_j' = \begin{cases} l, & \text{with probability } 1/2, \\ -l, & \text{with probability } 1/2. \end{cases}$$

Thus if $U_1 = 2l$, $\left(S_s\right)_{s=1}^{r}$ can be regarded as a symmetric random walk that starts from $2l$ in time step one, and then at each subsequent time step increases or decreases by l. Define T to be the first time the random walk hits 0, that is

$$T := \inf\left\{s \geq 2 : S_s = 0\right\} = \inf\left\{ s \geq 2 : \sum_{j=2}^{s}(U_j - U_j') = -2l \right\}.$$

By the reflection principle and symmetry,

$$\begin{aligned}
\mathbf{P}(T > r | U_1 = 2l) &= 1 - \mathbf{P}(T \leq r | U_1 = 2l) \\
&= 1 - \mathbf{P}\left(S_r = 0 | U_1 = 2l\right) - 2\mathbf{P}\left(S_r < 0 | U_1 = 2l\right) \\
&= 1 - \mathbf{P}\left(S_r = 0 | U_1 = 2l\right) - \mathbf{P}\left(S_r < 0 | U_1 = 2l\right) - \mathbf{P}\left(S_r > 4l | U_1 = 2l\right) \\
&= \sum_{k=1}^{4} \mathbf{P}\left(S_r = kl | U_1 = 2l\right) \leq 2 \max_{k \in Z} \mathbf{P}\left(S_r = kl | U_1 = 2l\right),
\end{aligned}$$

and by Lemma 4.7 of Barbour and Xia [1], we have

$$\max_{k\in Z} \mathbf{P}\left(S_r = kl | U_1 = 2l\right) \leq \frac{1}{\sqrt{2}} \frac{1}{\sqrt{r-1}},$$

thus

$$\mathbf{P}(T > r | U_1 = 2l) \leq \frac{2}{\sqrt{r}}. \tag{2}$$

Now for $j, s \in \{1, \ldots, r\}$ put

$$U_j'' := \begin{cases} U_j', & \text{if } 1 \leq j \leq T, \\ U_j, & j > T, \end{cases} \qquad \text{and} \quad V_s'' := \sum_{j=1}^{s} U_j''.$$

Of course $(V_s)_{s=1}^r$, $(V_s')_{s=1}^r$ and $(V_s'')_{s=1}^r$ all have the same distribution, thus $(V_r'', V_r + 1)$ is a coupling of $(V_r, V_r + 1)$, therefore

$$d_{TV}(V_r, V_r + 1) \leq \mathbf{P}(V_r + 1 \neq V_r'') = \mathbf{P}(T > r)$$

by the coupling inequality. Since

$$\mathbf{P}(T > r) = \mathbf{P}(U_1 = 2l)\mathbf{P}(T > r | U_1 = 2l) \leq \frac{1}{l\sqrt{r}}$$

by (2), the proof is complete. □

Now we show how the result of the lemma concerning sums of iid uniform random variables can be used to obtain similar results for sums of arbitrary integer valued random variables. The idea is to embed the uniform random variables in the ones we want to prove the result for.

Proposition. If $X_1, X_2, \ldots, X_n$, $n \geq 2$, are independent integer valued random variables and $W = \sum_{j=1}^n X_n$, then

$$d_{TV}(\mathscr{L}(W), \mathscr{L}(W+1)) \leq \frac{4}{l\sqrt{nlp}} + \frac{8d_n}{nlp},$$

where $l \in \{2, 4, 6, \ldots\}$ and $p \leq \min\{\mathbf{P}(X_j = k) : k = 1, \ldots, l, j = 1, \ldots, n\}$ are arbitrary and $d_n = d_{TV}(\mathscr{L}(X_n), \mathscr{L}(X_n + 1))$.

Proof. We write each of the variables $X_1, \ldots, X_n$ in the form

$$X_j = I_j U_j + (1 - I_j) R_j, \quad j = 1, \ldots, n, \tag{3}$$

where I_j, U_j and R_j, $j = 1, \ldots, n$, are all independent random variables defined on a common probability space, and for each $j = 1, \ldots, n$: U_j has discrete uniform distribution on $\{1, 2, \ldots, l\}$ for some even integer l; I_j is a Bernoulli random variable with parameter lp, where $p \leq \min\{\mathbf{P}(X_j = k) : k = 1, \ldots, l, j = 1, \ldots, n\}$ is fixed; and

$$\mathbf{P}(R_j = k) = \begin{cases} \frac{\mathbf{P}(X_j = k) - p}{1 - lp}, & 1 \leq k \leq l, \\ \frac{\mathbf{P}(X_j = k)}{1 - lp}, & \text{otherwise}, \end{cases} \qquad k \in Z.$$

Since $\mathscr{L}(X_j|I_j = 1) = \mathscr{L}(U_j)$ and $\mathscr{L}(X_j|I_j = 0) = \mathscr{L}(R_j)$, for any $\delta_1,\dots,\delta_{n-1} \in \{0,1\}$ and $\rho_1,\dots,\rho_{n-1} \in Z$ we have

$$\mathscr{L}\left(\sum_{j=1}^{n-1} X_j \Big| I_j = \delta_j, R_j = \rho_j, j = 1,\dots,n-1\right) = \mathscr{L}(V_r + \rho),$$

where $r = \sum_{j=1}^{n-1}\delta_j$, $V_r = \sum_{j=1}^{r} U_j'$, where the U_j' are independent copies of U_1, and are independent of everything else, and $\rho = \sum_{j=1}^{n-1}(1-\delta_j)\rho_j$.
Now we apply the inequality

$$d_{TV}(\mathscr{L}(Z_1),\mathscr{L}(Z_2)) \le \mathbf{E}\{d_{TV}(\mathscr{L}(Z_1|Z_3),\mathscr{L}(Z_2|Z_3))\} \tag{4}$$

true for any random elements Z_1, Z_2 and Z_3 defined on the same probability space. We obtain

$$\begin{aligned} & d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1)) \le \\ & \le \mathbf{E}\left\{ d_{TV}\left(\mathscr{L}\left(\sum_{j=1}^{n} X_j | I_j, R_j, j=1,\dots,n-1\right), \mathscr{L}\left(\sum_{j=1}^{n} X_j + 1 | I_j, R_j, j=1,\dots,n-1\right)\right)\right\} \\ & = \mathbf{E}\{d_{TV}(\mathscr{L}(V_T + X_n + R|T,R),\mathscr{L}(V_T + X_n + R + 1|T,R))\}, \end{aligned}$$

where $T = \sum_{j=1}^{n-1} I_j$ and $R = \sum_{j=1}^{n-1}(1-I_j)R_j$ are independent of $(U_j',\ j \ge 1)$ and of X_n. Hence

$$d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1)) \ \le\ \mathbf{E}\{d_{TV}(\mathscr{L}(V_T + X_n|T),\mathscr{L}(V_T + X_n + 1|T))\}, \tag{5}$$

since total variation distance is invariant under translation.
Now, since T, X_n and $(U_j',\ j \ge 1)$ are independent, we have

$$\begin{aligned} & d_{TV}(\mathscr{L}(V_T + X_n|T = t),\mathscr{L}(V_T + X_n + 1|T = t)) \\ & \qquad \le\ \min\{d_{TV}(\mathscr{L}(V_t),\mathscr{L}(V_t + 1)), d_{TV}(\mathscr{L}(X_n),\mathscr{L}(X_n + 1))\}, \end{aligned}$$

and the lemma provides the bound

$$d_{TV}(\mathscr{L}(V_t),\mathscr{L}(V_t + 1)) \le f(t) := \begin{cases} \frac{2}{l\sqrt{t}}, & \text{if } t > 0, \\ 1, & \text{if } t = 0. \end{cases}$$

Writing $d_n = d_{TV}(\mathscr{L}(X_n),\mathscr{L}(X_n + 1))$ we thus obtain from (5) that

$$\begin{aligned} d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1)) & \le \mathbf{E}\{d_{TV}(\mathscr{L}(V_T + X_n|T),\mathscr{L}(V_T + X_n + 1|T))\} \\ & \le \mathbf{E}\left\{\min\left[f(T); d_n\right]\right\} \\ & \le \mathbf{E}\left\{\frac{2}{l\sqrt{T}}\Big| T \ge \frac{\mathbf{E}T}{2}\right\}\mathbf{P}\left(T \ge \frac{\mathbf{E}T}{2}\right) + d_n\mathbf{P}\left(T < \frac{\mathbf{E}T}{2}\right) \\ & \le \frac{2\sqrt{2}}{l\sqrt{\mathbf{E}T}} + d_n\mathbf{P}\left(T < \frac{\mathbf{E}T}{2}\right). \end{aligned}$$

Since T has distribution $\mathrm{Bin}(n-1, lp)$, $\mathbf{E}T = (n-1)lp \ge \frac{1}{2}nlp$, and by Chebishev's inequality

$$\mathbf{P}\left(T < \frac{\mathbf{E}T}{2}\right) \le \mathbf{P}\left(|T - \mathbf{E}T| > \frac{\mathbf{E}T}{2}\right) \le \frac{4\mathbf{Var}T}{(\mathbf{E}T)^2} \le \frac{4}{(n-1)lp} \le \frac{8}{nlp},$$

thus

$$d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1)) \leq \frac{4}{l\sqrt{nlp}} + \frac{8d_n}{nlp}. \tag{6}$$

□

Remark 1. Since total variation distance is invariant under translation, there is no loss of generality in supposing that the l-intervals begin at 1.

Remark 2. The choice of (p,l) depends very much on the problem.

Remark 3. The constants in the upper bound of the proposition can be improved by refining the method proposed in the proof. One could embed not one, but many uniform random variables in the X_j-s by splitting the whole line into the l-blocks $(\{(m-1)l,\dots,ml\})_{m\in Z}$ and defining a uniform variable corresponding to each block. Thus one could use potential overlaps from the whole distribution and not just the interval $\{1,\dots,l\}$, when bounding $d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1))$. More precisely, each X_j, $j=1,\dots,n$, can be given in the form

$$X_j = I_{j0}R_j + \sum_{i=1}^{\infty} I_{ji}(U_{ji}+(i-1)l),$$

where all random variables in the decompositions are defined on a common probability space, and for each $j=1,\dots,n$ the following hold true: U_{ji} has discrete uniform distribution on $\{1,\dots,l\}$, $i=1,2,\dots$, for some fixed even integer l; $I_{j0} \sim$ Bernoulli $\left(1-\sum_{i=1}^{\infty} lp_i\right)$, $I_{ji} \sim$ Bernoulli(lp_i), where $p_i \leq \min\{\mathbf{P}(X_j=k) : k=(i-1)l,\dots,il, j=1,\dots,n\}$ is fixed, $i=1,2,\dots$, and these Bernoulli variables depend on each other in a way that for each outcome exactly one of them is 1 and the rest are 0; all the other variables in the decompositions are independent of each other and of the I_{ij}-s; and R_j is defined to make the distribution of the decomposition equal the distribution of X_j.
Then, to bound $d_{TV}(\mathscr{L}(W),\mathscr{L}(W+1))$ we would use (4), conditioning on all the I_{ji}-s and R_j-s, which would give us (5) with $T=\sum_{j=1}^{n-1}\sum_{i=1}^{\infty} I_{ji}$. In this case $\mathbf{E}T=(n-1)l\sum_{i=1}^{\infty} p_i$ and $\mathbf{Var}T=(n-1)l\left(\sum_{i=1}^{\infty} p_i\right)\left(1-l\sum_{i=1}^{\infty} p_i\right)$, hence we would obtain (6) with p replaced by $\sum_{i=1}^{\infty} p_i$.

Application to the coupon collector's problem

We return to the coupon collector's problem defined in the introduction, and our starting point is the well-known distributional equality ([5], p. 225)

$$W_{n,m} \stackrel{\mathscr{D}}{=} X_{m+1} + X_{m+2} + \dots + X_n, \tag{7}$$

where $X_{m+1}, X_{m+2}, \dots, X_n$ are independent random variables having geometric distributions with success probabilities $(m+1)/n, (m+2)/n, \dots, n/n$ respectively. Taking advantage of this decomposition we apply a theorem of Barbour and Xia [1] on translated compound Poisson approximation in total variation distance to the distributions of sums of independent integer valued random variables. One of the elements in their approximation error is (almost) $d_{TV}(\mathscr{L}(W_{n,m}),\mathscr{L}(W_{n,m}+1))$, to bound which we invoke our proposition. For $\mu, a>0$, define the compound Poisson distribution $\pi_{\mu,a}$ to be the distribution of Z_1+2Z_2, where $Z_1\sim\mathrm{Po}(\mu)$ and $Z_2\sim\mathrm{Po}(a/2)$ are independent. We have the following result:

Theorem. For any fixed $n \geq 2$ and $2 \leq m \leq n-4$, if

$$\begin{aligned} \mu &= \mathbf{Var}W_{n,m} - 2\langle \mathbf{Var}W_{n,m} - \mathbf{E}W_{n,m}\rangle, \\ a &= \langle \mathbf{Var}W_{n,m} - \mathbf{E}W_{n,m}\rangle \quad \text{and} \\ c &= \lfloor \mathbf{Var}W_{n,m} - \mathbf{E}W_{n,m}\rfloor, \end{aligned} \tag{8}$$

where $\langle x\rangle$ and $\lfloor x\rfloor$ denote the fractional and integer part of x respectively, then there exists a positive constant C such that

$$d_{TV}\left(\mathscr{L}\left(W_{n,m}+c\right),\pi_{\mu,a}\right) \leq C\frac{n}{m}\frac{1}{\sqrt{\mathbf{Var}W_{n,m}}}. \tag{9}$$

Remark 1. It has been proved by Baum and Billingsley in [6] that if $m = m_n \in \{0,1,\ldots,n-1\}$ is an integer that depends on n in such a way that

$$m_n \to \infty \quad \text{and} \quad \frac{n-m_n}{\sqrt{n}} \to \infty \quad \text{as} \quad n \to \infty, \tag{10}$$

then $\overline{W}_{n,m} := (W_{n,m_n} - \mathbf{E}W_{n,m_n})/\sqrt{\mathbf{Var}W_{n,m_n}}$ has asymptotically standard normal distribution. This limit theorem was refined in [7] by showing that

$$d_{Kol.}\left(\mathscr{L}\left(\overline{W}_{n,m}\right),\mathscr{L}(Z)\right) \leq C\frac{n}{m}\frac{1}{\sqrt{\mathbf{Var}W_{n,m}}}, \tag{11}$$

where $d_{Kol.}(\mathscr{L}(Z_1),\mathscr{L}(Z_2)) = \sup_{x\in R}|\mathbf{P}(Z_1 \leq x) - \mathbf{P}(Z_2 \leq x)|$ is the Kolmogorov distance, Z is a random variable of standard normal distribution and $C = 10.0245$, and that this order of approximation error is optimal. We see that the same order of approximation is obtained in the discrete approximation given in our theorem, but now with the error measured with respect to the much stronger total variation distance.

Remark 2. Note that, with these parameters, $\pi_{\mu,a}$ has mean $\mu+a = \mathbf{Var}W_{n,m} - \langle \mathbf{Var}W_{n,m} - \mathbf{E}W_{n,m}\rangle = \mathbf{E}W_{n,m} + c$ and variance $\mu + 2a = \mathbf{Var}W_{n,m}$.

Remark 3. We can express the bound above more intuitively with the help of the asymptotic formulae given by Baum and Billingsley in [6] for the variance of the waiting time: if $n \to \infty$, then

$$\begin{array}{lll} \frac{m}{n} \to 0, & \Rightarrow & \mathbf{Var}W_{n,m} \sim \frac{n^2}{m} \\ \frac{m}{n} \to c, c \in (0,1) & \Rightarrow & \mathbf{Var}W_{n,m} \sim \gamma n, \\ \frac{m}{n} \to 1, & \Rightarrow & \mathbf{Var}W_{n,m} \sim \frac{1}{2}\frac{(n-m)^2}{n}, \end{array}$$

where $\gamma = (1-c+c\log c)/c$. We shall refer to the categories above as "small", "medium" and "large" m. By these formulae, (9) is equivalent to

$$d_{TV}\left(\mathscr{L}\left(W_{n,m}+c_n\right),\pi_{\mu_n,a_n}\right) = \begin{cases} O\left(\frac{1}{\sqrt{m}}\right), & \text{in the "small" } m \text{ case;} \\ O\left(\frac{1}{\sqrt{n}}\right), & \text{in the "medium" } m \text{ case;} \\ O\left(\frac{\sqrt{n}}{n-m}\right), & \text{in the "large" } m \text{ case.} \end{cases} \tag{12}$$

Proof. We apply Theorem 4.3 in [1], which states that if Z_j, $j = 1,\dots,r$, are independent integer valued random variables with $\mathbf{E}|Z_j|^3 < \infty$, $W = \sum_{j=1}^r Z_j$, and we define

$$\psi_j := \mathbf{E}|Z_j(Z_j-1)(Z_j-2)| + |\mathbf{E}Z_j|\mathbf{E}|Z_j(Z_j-1)| + 2\mathbf{E}|Z_j||\mathbf{Var}Z_j - \mathbf{E}Z_j|,$$

$$d_+ := \max_{1\le i\le r}\left\{d_{TV}(\mathscr{L}(W_i),\mathscr{L}(W_i+1))\right\}, \text{ where } W_i := W - Z_i,$$

then with $\mu = \mathbf{Var}W - 2\langle \mathbf{Var}W - \mathbf{E}W\rangle$, $a = \langle \mathbf{Var}W - \mathbf{E}W\rangle$ and $c = \lfloor \mathbf{Var}W - \mathbf{E}W\rfloor$,

$$d_{TV}\left(\mathscr{L}(W+c),\pi_{\mu,a}\right) \le \frac{2 + 2\left(|\lfloor \mathbf{Var}W - \mathbf{E}W\rfloor| + \sum_{j=1}^r \psi_j\right) d_+}{\mathbf{Var}W}. \tag{13}$$

We apply this theorem with $Z_j = X_j - 1$, $j \in \{m+1,\dots,n\}$, for the X_j given in (7), in order to approximate the coupon collector's shifted waiting time $\widetilde{W}_{n,m} := \sum_{j=m+1}^n [X_j - 1]$, and then show that the upper bound in (13) is not greater than the right hand side of (9). Then, since the two measures compared in (13) are the same for $W = \widetilde{W}_{n,m}$ and $W = W_{n,m} = \widetilde{W}_{n,m} + n - m$, the theorem for $W_{n,m}$ follows immediately.
To do so, for given $n \ge 2$, $2 \le m \le n-4$ and $j \in \{m+1,\dots,n\}$, we bound ψ_j and d_+ as defined above. For X, a random variable that has geometric distribution with parameter p, we have

$$\mathbf{E}X = \frac{1}{p}, \quad \mathbf{E}X^2 = \frac{2-p}{p^2}, \quad \mathbf{E}X^3 = \frac{p^2-6p+6}{p^3}, \text{ and } \mathbf{Var}X = \frac{1-p}{p^2}.$$

Hence for $Z = X - 1$, one can easily calculate

$$\begin{aligned}\psi :&= \mathbf{E}|Z(Z-1)(Z-2)| + |\mathbf{E}Z|\mathbf{E}|Z(Z-1)| + 2\mathbf{E}|Z||\mathbf{Var}Z - \mathbf{E}Z| \\ &= \mathbf{E}\{X^3 - 6X^2 + 11X - 6\} + \mathbf{E}\{X-1\}\mathbf{E}\{X^2 - 3X + 2\} + 2\mathbf{E}\{X-1\}|\mathbf{Var}X - \mathbf{E}X + 1| \\ &= \frac{10(1-p)^3}{p^3},\end{aligned}$$

so $\psi_j = 10\left(\frac{n}{j} - 1\right)^3$. If we add the ψ_j together, we obtain

$$\sum_{j=m+1}^n \psi_j = \sum_{j=m+1}^n \frac{10\left(1-\frac{j}{n}\right)^3}{\left(\frac{j}{n}\right)^3} \le 10\frac{n}{m+1}\sum_{j=m+1}^n \frac{1-\frac{j}{n}}{\left(\frac{j}{n}\right)^2} = 10\frac{n}{m+1}\mathbf{Var}W_{n,m}.$$

Now combining the bound above with inequality (13) applied to the $\widetilde{W}_{n,m}$ waiting time, and also noticing that $0 \le \mathbf{Var}W_{n,m} - \mathbf{E}\widetilde{W}_{n,m} \le \mathbf{Var}W_{n,m}$, because for each X_j geometric random variable of parameter j/n we have $\mathbf{Var}X_j - \mathbf{E}(X_j - 1) = \left(\frac{1-j/n}{j/n}\right)^2 \le \frac{1-j/n}{(j/n)^2} = \mathbf{Var}X_j$, we obtain

$$d_{TV}\left(\mathscr{L}\left(\widetilde{W}_{n,m} + \tilde{c}\right),\pi_{\tilde{\mu},\tilde{a}}\right) \le \frac{2}{\mathbf{Var}W_{n,m}} + \left(2 + \frac{20n}{m+1}\right)d_+, \tag{14}$$

where $\tilde{c}$, $\tilde{\mu}$ and $\tilde{a}$ are defined by the formulae in (8) with $W_{n,m}$ replaced with $\widetilde{W}_{n,m}$.
Before turning to the approximation of d_+, we bound $\mathbf{Var}W_{n,m}$. We see that

$$\mathbf{Var}W_{n,m} = n\sum_{j=m+1}^n \frac{n-j}{j^2} \le \frac{n}{(m+1)^2}\sum_{j=m+1}^n (n-j) = \frac{n(n-m)(n-m-1)}{2(m+1)^2},$$

also,

$$\mathbf{Var}W_{n,m} = n\sum_{j=m+1}^{n}\frac{n-j}{j^2} \le n(n-m-1)\int_m^n \frac{1}{x^2}dx \le \frac{n(n-m-1)}{m},$$

thus

$$\mathbf{Var}W_{n,m} \le n(n-m-1)\min\left\{\frac{n-m}{2(m+1)^2}, \frac{1}{m}\right\}. \tag{15}$$

Now for d_+, by an inequality of Mattner and Roos [4] we have

$$d_{TV}(\mathscr{L}(W_i), \mathscr{L}(W_i+1)) \le \sqrt{\frac{2}{\pi}}\left(\sum_{j=m+1, j\neq i}^{n}\left[1-d_{TV}(\mathscr{L}(X_j), \mathscr{L}(X_j+1))\right]\right)^{-\frac{1}{2}},$$

and since $d_{TV}(\mathscr{L}(X_j), \mathscr{L}(X_j+1))$ is equal to

$$\frac{1}{2}\sum_{k=1}^{\infty}|\mathbf{P}(X_j=k)-\mathbf{P}(X_j=k-1)| = \frac{1}{2}\left(\frac{j}{n}+\left(\frac{j}{n}\right)^2\sum_{k=2}^{\infty}\left(1-\frac{j}{n}\right)^{k-2}\right) = \frac{j}{n}, \tag{16}$$

we obtain

$$d_+ \le \sqrt{\frac{2}{\pi}}\left(\sum_{j=m+1}^{n}\left(1-\frac{j}{n}\right) - \max_{m+1\le i\le n}\left(1-\frac{i}{n}\right)\right)^{-\frac{1}{2}} = \sqrt{\frac{2}{\pi}}\frac{\sqrt{n}}{\sqrt{(n-m-1)(n-m-2)}}.$$

It follows from this and (15) that for any $K>0$

$$d_+ \le \frac{2K}{\sqrt{\pi}}\frac{1}{\sqrt{\mathbf{Var}W_{n,m}}}, \quad \text{if } \frac{n}{K} \le m \le n-4. \tag{17}$$

Putting this bound into (14) gives a result which, when compared to (9), has an extra factor $K \ge n/m$. Thus it is of inferior order if $m \ll n$. To prove the theorem for such values of m, we need to use our proposition to bound d_+.

Let us assume that $2 \le m \le \frac{n}{2}$. If we apply the Proposition to the random variables $\{X_j, j = m+1, \dots, 2m, j \neq i\}$, $i \in \{m+1, \dots, 2m\}$ fixed, with

$$l = \begin{cases} \lfloor \frac{n}{m} \rfloor, & \text{if } \lfloor \frac{n}{m} \rfloor \text{ is even,} \\ \lfloor \frac{n}{m} \rfloor - 1, & \text{if } \lfloor \frac{n}{m} \rfloor \text{ is odd,} \end{cases} \quad \text{and} \quad p = \left(1-\frac{2m}{n}\right)^l \frac{m}{n},$$

we obtain

$$d_{TV}\left\{\mathscr{L}\left(\sum_{j=m+1, j\neq i}^{2m} X_j\right), \mathscr{L}\left(\sum_{j=m+1, j\neq i}^{2m} X_j + 1\right)\right\} \le \frac{2}{l\sqrt{(m-1)lp}} + \frac{8d}{(m-1)lp}, \tag{18}$$

where

$$d = \begin{cases} d_{TV}\left\{\mathscr{L}(X_{2m}), \mathscr{L}(X_{2m}+1)\right\} = \frac{2m}{n}, & \text{if } i \neq 2m, \\ d_{TV}\left\{\mathscr{L}(X_{2m-1}), \mathscr{L}(X_{2m-1}+1)\right\} = \frac{2m-1}{n}, & \text{if } i = 2m \end{cases}$$

by (16). For any $i \in \{m+1, \dots, 2m\}$ we have

$$d \le \frac{2m}{n} \quad \text{and} \quad l \ge \frac{n}{2m},$$

since $\lfloor x \rfloor - 1 \geq \frac{x}{2}$, if $x \geq 2$, and

$$lp \geq \frac{n}{2m}\left(1-\frac{2m}{n}\right)^{\frac{n}{m}}\frac{m}{n} \geq \frac{e^{-2}}{2},$$

because $(1-x)^{\frac{2}{x}}$ decreases as x increases in (0,1), and its limit at 0 is e^{-2}. Now putting the bounds above together in (18) yields

$$\begin{aligned} d_+ &= \max_{i\in\{m+1,\ldots,2m\}} d_{TV}\left\{\mathscr{L}\left(\sum_{j=m+1,j\neq i}^{2m} X_j\right), \mathscr{L}\left(\sum_{j=m+1,j\neq i}^{2m} X_j+1\right)\right\} \\ &\leq 8\sqrt{2}e\frac{m}{\sqrt{m-1}n} + 32e^2\frac{m}{(m-1)n} \leq 16e\frac{\sqrt{m}}{n} + 64e^2\frac{1}{n} \leq (16e+64e^2)\frac{\sqrt{m}}{n}, \end{aligned}$$

where the last two inequalities hold for $m \geq 2$. By (15), $\frac{\sqrt{m}}{n} \leq \frac{1}{\sqrt{\mathbf{Var}W_{n,m}}}$, thus we have

$$d_+ \leq (16e+64e^2)\frac{1}{\sqrt{\mathbf{Var}W_{n,m}}}, \quad \text{if } 2 \leq m \leq \frac{n}{2}.$$

This and (17) with $K=2$ substituted into (14) yield the theorem. □

Acknowledgements. I would like to thank Andrew Barbour for introducing me to the problem and for his invaluable help. I am also grateful to the Institute for Mathematical Sciences of the National University of Singapore for their hospitality, part of the work was done during a visit there.

References

[1] Barbour, A. D. and Xia, A., Poisson Perturbations, *ESAIM Probab. and Statist.* **3** (1999), 131–150. Math. Review 2000j:60026 MR1716120

[2] Barbour, A. D. and Cekanavicius, V, Total variation asymptotics for sums of independent integer random variables, *The Annals of Probability* **30** (2002), 509–545. Math. Review 2003g:60072 MR1905850

[3] Lindvall, T., *Lectures on the Coupling Method*, Dover Publications, Mineola, NY, 1992. MR1180522

[4] Mattner, L. and Roos, B., A shorter proof of Kanter's Bessel function concentration bound, *Probability Theory and Related Fields* **139** (2006), 191–205. Math. Review 2008e:60044 MR2322695

[5] Feller, W., *An Introduction to Probability Theory and its Applications*, John Wiley & Sons, 1968. MR0228020

[6] Baum, L. E. and Billingsley, P., Asymptotic distributions for the coupon collector's problem, *Ann. Math. Statist.* **36** (1965), 1835–1839. Math. Review 31 #6263 MR0182039

[7] Pósfai, A., Rates of convergence for the normal approximation in incomplete coupon collection, *Acta. Sci. Math. (Szeged)* **73** (2007), 333-348. Math. Review 2008j:60062 MR2339869