Topic 7
Moments of Randomly Stopped Sums of Independent Variables

Victor H. de la Peña

Professor of Statistics, Columbia University

Artificial Intelligence Institute for Advances in Optimization
Georgia Institute of Technology 2024

Wald's Identity

We begin with Wald's equations, which constitute the cornerstone of the theory of sequential analysis.

Theorem (Wald, 1944 [5])

Let X_i be a sequence of i.i.d. random variables adapted to $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$, with $\mathbb{E}(X) = \mu$, $|\mu| < \infty$. Let T be a stopping time adapted to $\sigma(X_i)$. Set $S_n = X_1, + \dots + X_n$. Then

$$\mathbb{E}(S_T) = \mu \mathbb{E}(T), \text{ whenever } \mathbb{E}(T) < \infty \tag{1}$$

Moreover, if $\mathbb{E}(X_1) = 0$, and $\mathbb{E}({X_1}^2) < \infty$, then

$$\mathbb{E}(S_T^2) = \mathbb{E}({X_1}^2)\mathbb{E}(T), \text{ whenever } \mathbb{E}(T) < \infty. \tag{2}$$

In this context, although the stopping time T is adapted to $\mathbb{F} := \{\mathcal{F}_i\}_{i\geq 1} := \{\sigma(X_1, ..., X_i)\}_{i\geq 1}$, we can decouple this structure and still keep the identity, the proof of which is even simpler than the Wald identity.

Theorem

Let X_i be a sequence of independent random variables adapted to $\mathcal{F}_i = \sigma(X_1, ..., X_i)$, with $\mathbb{E}(X_i) = \mu_i$, $|\mu_i| < \infty$ for all i. Let T be a stopping time adapted to $\mathbb{F}$. Let $\{\tilde{X}_i\}$ be the i.i.d. copy of $\{X_i\}$, and be independent of T. Set $S_n = X_1, + ... + X_n$, and $\tilde{S}_n = \tilde{X}_1, + ... + \tilde{X}_n$. Then

$$\mathbb{E}(S_T) = \mathbb{E}(\tilde{S}_T), \textit{ whenever } \mathbb{E}(T) < \infty \tag{3}$$

Moreover, if $\mathbb{E}(X_i) = 0$, and $\mathbb{E}({X_1}^2) < \infty$, then

$$\mathbb{E}(S_T^2) = \mathbb{E}(\tilde{S}_T^2), \textit{ whenever } \mathbb{E}(T) < \infty. \tag{4}$$

Bounding $\mathbb{E}S_T^2$ via $\mathbb{E}\tilde{S}_T^2$

We now consider the second moment of S_T, where X_i's are square-integrable but may not be mean-zero. Although the equation (4) for the mean-zero random variables no longer holds, we may still use the second moment of $\tilde{S}_T$ to bound $\mathbb{E}S_T^2$. In the following discussion, we presume that X_i's are independent (but may not be identically distributed) and square-integrable, and $\mathbb{E}T < \infty$.

Remark: We note that $\tilde{S}_T$ has the same distribution of $S_{\tilde{T}}$, where $\tilde{T}$ is an independent copy of T.

Lemma

Let $\mathbb{E}X_i = \mu_i$ and $\mathbb{V}ar(X_i) = \sigma_i^2$, and let $\tilde{T}$ an independent copy of T. Then

$$\mathbb{E}S_{\tilde{T}}^2 = \mathbb{E}\sum_{i=1}^{\tilde{T}} \sigma_i^2 + \mathbb{E}\left(\sum_{i=1}^{\tilde{T}} \mu_i\right)^2 = \mathbb{E}\sum_{i=1}^{T} \sigma_i^2 + \mathbb{E}\left(\sum_{i=1}^{T} \mu_i\right)^2. \tag{5}$$

$$\begin{aligned}
\mathbb{E}S_{\tilde{T}}^2 &= \mathbb{E}\left(\sum_{i=1}^{\tilde{T}}(X_i - \mu_i) + \sum_{i=1}^{\tilde{T}} \mu_i\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{\tilde{T}}(X_i - \mu_i)\right)^2 + 2\mathbb{E}\left(\sum_{i=1}^{\tilde{T}}(X_i - \mu_i)\right)\left(\sum_{i=1}^{\tilde{T}} \mu_i\right) + \mathbb{E}\left(\sum_{i=1}^{\tilde{T}} \mu_i\right)^2 \\
&= \mathbb{E}\sum_{i=1}^{\tilde{T}} \sigma_i^2 + \mathbb{E}\left(\sum_{i=1}^{\tilde{T}} \mu_i\right)^2,
\end{aligned}$$

where we establish the last equation by conditioning $\tilde{T}$, and (5) holds due to (3) and (4).

Lemma

$$\mathbb{E}\left(\sum_{i=1}^{T}\mu_i\right)^2 = \mathbb{E}S_T^2 + \mathbb{E}\left(\sum_{i=1}^{T}(X_i-\mu_i)\right)^2 - 2\mathbb{E}\left(S_T\sum_{i=1}^{T}(X_i-\mu_i)\right)$$

$$\leq \left(\sqrt{\mathbb{E}S_T^2} + \sqrt{\mathbb{E}\sum_{i=1}^{T}\sigma_i^2}\right)^2, \tag{6}$$

$$\mathbb{E}S_T^2 = \mathbb{E}\left(\sum_{i=1}^{T}\mu_i\right)^2 + \mathbb{E}\left(\sum_{i=1}^{T}(X_i-\mu_i)\right)^2 - 2\mathbb{E}\left(\sum_{i=1}^{T}\mu_i\sum_{i=1}^{T}(X_i-\mu_i)\right)$$

$$\leq \left(\sqrt{\mathbb{E}\left(\sum_{i=1}^{T}\mu_i\right)^2} + \sqrt{\mathbb{E}\sum_{i=1}^{T}\sigma_i^2}\right)^2. \tag{7}$$

Both first equations are obtained directly from $\sum_{i=1}^{T}\mu_i = S_T - \sum_{i=1}^{T}(X_i-\mu_i)$, and both second inequalities are due to the Cauchy-Schwarz inequality and completing the square.

From the above two lemmas, we have that

$$\mathbb{E}S_T^2 \leq \left(\sqrt{\mathbb{E}\left(\sum_{i=1}^{T} \mu_i\right)^2} + \sqrt{\mathbb{E}\sum_{i=1}^{T} \sigma_i^2} \right)^2$$
$$\leq 2\left[\mathbb{E}\left(\sum_{i=1}^{T} \mu_i\right)^2 + \mathbb{E}\sum_{i=1}^{T} \sigma_i^2\right] = 2\mathbb{E}S_{\tilde{T}}^2.$$

Note that $S_{\tilde{T}}$ and $\tilde{S}_T$ have the same distributions, the following inequality is induced.

Theorem (de la Peña & Govindarajulu, 1992 [4])

$$0 \leq \mathbb{E}S_T^2 \leq 2\mathbb{E}\tilde{S}_T^2. \tag{8}$$

In addition, since almost surely $\sum_{i=1}^{T} \mathbb{E} X_i^2 \geq \sum_{i=1}^{T} \sigma_i^2$,

$$\begin{aligned}\mathbb{E}\sum_{i=1}^{T} X_i^2 &= \sum_{i=1}^{\infty} \mathbb{E}(X_i^2 I_{T\geq i}) = \sum_{i=1}^{\infty}[\mathbb{E}(X_i^2)\mathbb{E} I_{T\geq i}] \\ &= \mathbb{E}[(\sum_{i=1}^{\infty} \mathbb{E}(X_i^2) I_{T\geq i}] = \mathbb{E}\sum_{i=1}^{T} \mathbb{E} X_i^2.\end{aligned}$$

When all X_i's are non-negative, we have $\mathbb{E} S_T^2 \geq \mathbb{E}\sum_{i=1}^{T} X_i^2 \geq \mathbb{E}\sum_{i=1}^{T} \sigma_i^2$. Then from (6) we have

$$\mathbb{E}\left(\sum_{i=1}^{T} \mu_i\right)^2 \leq \left(\sqrt{\mathbb{E} S_T^2} + \sqrt{\mathbb{E}\sum_{i=1}^{T} \sigma_i^2}\right)^2 \leq 4\mathbb{E} S_T^2. \qquad (9)$$

Therefore, we have

$$\mathbb{E}S_{\tilde{T}}^2 = \mathbb{E}\left(\sum_{i=1}^{T}\mu_i\right)^2 + \mathbb{E}\sum_{i=1}^{T}\sigma_i^2 \leq 5\mathbb{E}S_T^2,$$

which leads to the following theorem.

Theorem (de la Peña & Govindarajulu, 1992 [4])

With the same assumption above, we further suppose that for all $i = 1, \ldots, n$, $X_i \geq 0$ *almost surely, then*

$$\mathbb{E}S_T^2 \geq \frac{1}{5}\mathbb{E}S_{\tilde{T}}^2. \tag{10}$$

We further note that the bounds $0 \leq \mathbb{E}S_T^2 \leq 2\mathbb{E}\tilde{S}_T^2$ is sharp, from the following example, provided by Aryeh Dvoretzky.
Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with

$$Y_1 = \begin{cases} 1, \; w.p. \dfrac{1}{n} \\ \dfrac{-1}{n-1}, \; w.p. \dfrac{n-1}{n} \end{cases}$$

Then $\mathbb{E}Y_1 = 0$ and $\mathbb{E}Y_1^2 = (n-1)^{-1}$. Let the stopping time

$$T_n = \begin{cases} 1, \text{ if } Y_1 < 0 \\ k_n, \text{ if } Y_1 > 0 \end{cases},$$

for some k_n such that $k_n/n \to 0$ and $k_n^2/n \to \infty$ when $n \to \infty$.
In this case, marginally, when $n \to \infty$, $\mathbb{E}T_n = 1 - n^{-1} + \frac{k_n}{n} \to 1$, and $\mathbb{E}T_n^2 = 1 - n^{-1} + \frac{k_n^2}{n} \sim \frac{k_n^2}{n}$.

Setting some constant $a \in \mathbb{R}$ and we then have

$$\mathbb{E}\left(\sum_{i=1}^{T_n}(a+Y_i)\right)^2 = a^2\mathbb{E}T_n^2 + 2a\mathbb{E}[T_n\sum_{i=1}^{T_n}Y_i] + \mathbb{E}[\sum_{i=1}^{T_n}Y_i]^2$$
$$\sim a^2\frac{k_n^2}{n} + 2a\frac{k_n}{n} + \frac{1}{n}.$$

When we let $a = \frac{1}{k_n}$, $\mathbb{E}\left(\sum_{i=1}^{T_n}(a+Y_i)\right)^2 \sim \frac{4}{n}$. And when $a = \frac{-1}{k_n}$, $\mathbb{E}\left(\sum_{i=1}^{T_n}(a+Y_i)\right)^2 = o(n^{-1})$. By comparison, for the i.i.d. copy $\tilde{T}_n$ of T_n, we also have when $a = \pm\frac{1}{k_n}$

$$\mathbb{E}\left(\sum_{i=1}^{\tilde{T}_n}(a+Y_i)\right)^2 = \mathbb{E}\tilde{T}_n\mathbb{E}Y_1^2 + a^2\mathbb{E}\tilde{T}_n^2 \sim \frac{1}{n} + \frac{a^2k_n^2}{n} \sim \frac{2}{n}.$$

Hence, both the upper bound and the lower bound are sharp.

Application

Consider the hitting time $T_r := \inf\{n : S_n^2 \geq r\}$ for some nonnegative r, and the function $a : \mathbb{N}_0 \to \mathbb{R}_0^+$ induced by S_n such that

$$a(n) := \mathbb{E}\left[\max_{0 \leq j \leq n} S_j^2\right], \quad \forall n \in \mathbb{N}_0.$$

Then we can lower bound the expectation of the random variable $a(T_r)$, via the following procedure

$$\begin{aligned} & r \leq \mathbb{E}[S_{T_r}^2] \leq 2\mathbb{E}[\tilde{S}_{T_r}^2] \leq 2\mathbb{E}\left[\max_{0 \leq j \leq T_r} \tilde{S}_j^2\right] \\ \iff & \mathbb{E}[a(T_r)] \geq r/2. \end{aligned}$$

Remark: This result can be extended to the case for all nonnegative, measurable process X_t with $a(t) = \mathbb{E} \sup_{0 \le s \le t} X_s$ and $T_r := \inf\{t : X_t \ge r\}$, such that $r/2 \le \mathbb{E}[a(T_r)]$ (see Brown, de la Peña & Sit [1]).
If $a(t)$ is assumed to be concave, we obtain that

$$a^{-1}(T_r) \le \mathbb{E}[T_r]. \tag{11}$$

If $a(\cdot)$ is continuous and strictly increasing, there is a sharp inequality for any Cadlag stochastic process, X_t with $X_0 = 0$ and $g(\cdot)$ non-decreasing (see Brown, de la Peña, Klass & Sit [2])

$$\mathbb{E}g(T_r) \ge \int_0^1 g\left(a^{-1}(r\alpha)\right) d\alpha. \tag{12}$$

Extension

In 1993, Hitczenko [3] extended the inequality to p-th moment.

Theorem

With the same assumptions above, we further assume that for all $i = 1, 2..., X_i \geq 0$ *almost surely, then for all* $1 \leq p < \infty$,

$$\mathbb{E}S_T^p \leq 2^{p-1}\mathbb{E}\tilde{S}_T^p. \tag{13}$$

Remark 1: This bound is proved to be sharp.
Remark 2: This bound is established through a more general result in tangent decoupling. We will introduce them soon.

References I

[1] Mark Brown, Victor de la Peña, and Tony Sit. "From boundary crossing of non-random functions to boundary crossing of stochastic processes". In: *Probability in the Engineering and Informational Sciences* 29.3 (2015), pp. 345–359.

[2] Mark Brown et al. "On an approach to boundary crossing by stochastic processes". In: *Stochastic Processes and their Applications* 126.12 (2016), pp. 3843–3853.

[3] Pawel Hitczenko. "Sharp inequality for randomly stopped sums of independent non-negative random variables". In: *Stochastic processes and their applications* 51.1 (1994), pp. 63–73.

[4] Victor H de la Peña and Z Govindarajulu. "A note on second moment of a randomly stopped sum of independent variables". In: *Statistics & Probability Letters* 14.4 (1992), pp. 275–281.

References II

[5] Abraham Wald. "On cumulative sums of random variables". In: *The Annals of Mathematical Statistics* 15.3 (1944), pp. 283–296.