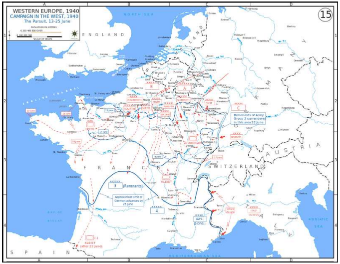
Generalizing the German Tank Problem: Math/Stats at War!



Anthony Lee: Milton Academy
(anthony_lee24@milton.edu)
Advisor: Steven J. Miller: Williams College
(sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu)
http://www.williams.edu/Mathematics/sjmiller

AISC Conference at UNC Greensboro, October 8, 2022

Introduction

The German Tank Problem: General Statement

- Problem of statistical inference using limited information.
- Dates back to WW II.
- Allies needed to find an accurate way of estimating the number of German Tanks being produced.

Approaches

How did allies find the number of tanks?

Approaches

How did allies find the number of tanks?

Spies vs Math/Stats!

Month	Statistical estimate	Intelligence estimate	German records
June 1940	169	1,000	122
June 1941	244	1,550	271
August 1942	327	1,550	342

Allies used serial numbers of captured or broken tanks on the battlefield to find a formula for estimation.

Assumed serial numbers started at 1.

Statistical estimates were MUCH more accurate.

Importance

Why is it important to make a accurate estimate?

Dangers of under-estimating and over-estimating.

Importance

Why is it important to make a accurate estimate?

Dangers of under-estimating and over-estimating.

- Underestimating: Send too few tanks and will likely lose in battle.
- Overestimating: Waste resources, waste time.

Structure of the Talk

- Original Tank Problem & Improving the formula.
- Discrete 2-dim Problem (Square & Circle).
- Discrete Generalized L dimensional problem.

Mathematical Preliminaries

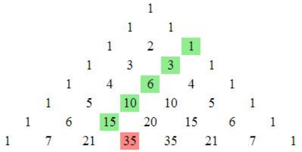
Combinatorial Identities

Pascal's identity

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

Hockey Stick Identity

$$\sum_{m=k}^{N} \binom{m}{k} = \binom{N+1}{k+1}.$$



Generalized Hockey Stick Identities

Identity I: For all $N \geq k$,

$$\sum_{m=k}^{N} \binom{m-b}{k-c} \;=\; \binom{N-b+1}{k-c+1} - \binom{k-b}{k-c+1}.$$

Identity II: For all $N \geq k$,

$$\sum_{m=k}^{N} (m-a) \frac{\binom{m-b}{k-c}}{\binom{N}{k}} \;=\; \frac{N\binom{N-b+1}{k-c+1} - (k-1)\binom{k-b}{k-c+1} - \binom{N-b+1}{k-c+2}}{\binom{N}{k}} - \frac{a\binom{N-b+1}{k-c+1} - \binom{k-b}{k-c+2} - a\binom{k-b}{k-c+1}}{\binom{N}{k}}.$$

CDF Method

Cumulative Distribution Function Method

The cumulative distribution function (CDF) of a random variable X with density f, denoted F, is given by

$$F(x) \ := \ \text{Prob}(X \le n) \ = \ \int_{-\infty}^{n} f(t)\,dt, \ \text{for any } n \in \mathbb{R}.$$

Discrete CDF method: get probability by differences

$$\text{Prob}(X = n) \ = \ \text{Prob}(X \le n) \ - \ \text{Prob}(X \le n-1).$$

Variance & Covariance

Definition

The variance for a random variable X is the average of the squared difference from the mean, $\mathbb{E}[X]$:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2. \end{aligned}$$

Lemma

The covariance of two random variables satisfies

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathbb{E}[X - \mathbb{E}[X]] \cdot \mathbb{E}[Y - \mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]. \end{aligned}$$

1-dim German Tank Problem

Conjecturing the Formula

How should $\widehat{N}$ depend on m_k (maximum tank number observed) and k (number of tanks observed)?

Conjecturing the Formula

How should $\widehat{N}$ depend on m_k (maximum tank number observed) and k (number of tanks observed)?

We state the formula:

$$\widehat{N} \;=\; m_k\left(1+\frac{1}{k}\right)-1.$$

Is this reasonable?

Conjecturing the Formula

How should $\widehat{N}$ depend on m_k (maximum tank number observed) and k (number of tanks observed)?

We state the formula:

$$\widehat{N} \;=\; m_k\left(1+\frac{1}{k}\right)-1.$$

Is this reasonable?

Note sanity checks at $k=1$ and $k=N$.

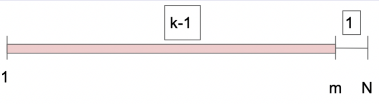
Proving the Formula: Step 1: Sample Maximum Probability

Lemma

M: random variable for the maximum number observed; m_k: the value we see. For $k \leq m_k \leq N$,

$$\text{Prob}(M = m_k) \;=\; \frac{\binom{m_k - 1}{k-1}}{\binom{N}{k}}.$$

Proof:



Have to choose $(k-1)$ tanks from $(m-1)$ values. Probability is $\binom{m-1}{k-1} / \binom{N}{k}$. $\square$

Step 2: Expected Value Preliminaries

Definition of expected value:

$$\mathbb{E}[M] \ := \ \sum_{m_k=k}^{N} m_k \cdot \mathrm{Prob}(M = m_k).$$

Substituting:

$$\mathbb{E}[M] \ = \ \sum_{m_k=k}^{N} m_k \frac{\binom{m_k-1}{k-1}}{\binom{N}{k}}.$$

Step 3: Computing $\mathbb{E}[M]$

$$\mathbb{E}[M] = \sum_{m_k=k}^{N} m_k \frac{\binom{m_k-1}{k-1}}{\binom{N}{k}}.$$

Hockey stick identity allows us calculate nice closed form expressions.

$$\mathbb{E}[M] = \frac{k(N+1)}{k+1}.$$

We invert the equation.

Step 4: Finding $\widehat{N}$

Solve for N:

$$N = \mathbb{E}[M]\left(1 + \frac{1}{k}\right) - 1.$$

Substitute m_k (observed value for M) as best guess for $\mathbb{E}[M]$.

Obtain our estimate for the number of tanks produced:

$$\widehat{N} = m_k\left(1 + \frac{1}{k}\right) - 1,$$

completing the proof. $\square$

Improving the formula

Can we do better by using the second largest tank? What about the L^{th} largest tank?

Improving the formula

Can we do better by using the second largest tank? What about the L^{th} largest tank?

Probability that the second largest tank is m_{k-1}:

$$\text{Prob}(M_{k-1} = m_{k-1}) \;=\; \frac{\binom{m_{k-1}-1}{k-2}\binom{N-m_{k-1}}{1}}{\binom{N}{k}}.$$

Formula using Second Largest Tank

Using these identities, we get:

$$\widehat{N} \;=\; m_{k-1}\frac{k+1}{k-1} \;-\; 1.$$

Using the definition of variance, $\mathbb{E}[X^2] \;-\; \mathbb{E}[X]^2$, to calculate variances of both formulas.

$$\mathrm{Var}(X_k) \;=\; \frac{(N-k)(N+1)}{(k)(k+2)}.$$

$$\mathrm{Var}(X_{k-1}) \;=\; \frac{2(N-k)(N+1)}{(k+2)(k-1)}.$$

Notice that $\mathrm{Var}(X_{k-1})$ is roughly two times as larger.

Using more tanks to estimate

What if we use more than one tank?

Using more tanks to estimate

What if we use more than one tank?

Motivation from portfolio theory.

Linear combination of two stocks.

Two independent stocks with same rate of return with different variances.

Can get a smaller variance by investing in both.

Portfolio Theory

Two independent stocks with mean μ, variances σ_i (assume $0 \le \sigma_2 \le \sigma_1$).

Let $X_\alpha = \alpha X_1 + (1-\alpha)X_2$. We have

$$\mathrm{Var}(X_\alpha) \;=\; \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2.$$

Check endpoints ($\alpha = 0$ or 1) and critical points:

$$2\alpha\sigma_1^2 - 2(1-\alpha)\sigma_2^2 \;=\; 0,$$

gives a critical value of

$$\alpha_* \;=\; \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Checking Critical Values

- When $\alpha = 0$, $\text{Var}(X_\alpha) = \sigma_2^2$.
- When $\alpha = 1$, $\text{Var}(X_\alpha) = \sigma_1^2$.
- When $\alpha = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$, plug in α:

$$\begin{aligned}\text{Var}(X_\alpha) &= \left(\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^2 \sigma_1^2 + \left(1 - \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^2 \sigma_2^2 \\ &= \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}.\end{aligned}$$

Notice

$$\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} \cdot \sigma_2^2 < \sigma_2^2.$$

Weighted Statistic

We quickly review notation.

- M_k = Largest Observed Tank.
- M_{k-1} = Second largest Observed Tank.
- X_k = Statistic using M_k: $M_k(\frac{k+1}{k})$ - 1.
- X_{k-1} = Statistic using M_{k-1}: $M_{k-1}(\frac{k+1}{k-1})$ - 1.

Weighted statistic. Let $\alpha \in [0, 1]$ and define the weighted statistic X_α by

$$X_\alpha \ := \ \alpha X_k \ + \ (1-\alpha)X_{k-1}.$$

Variance of Weighted Statistic

We calculate the variance of X_α.

$$\begin{aligned}\mathrm{Var}(X_\alpha) &= \alpha^2\mathrm{Var}(X_k) + (1-\alpha)^2\mathrm{Var}(X_{k-1}) \\ &\quad + 2\alpha(1-\alpha)\mathrm{Cov}(X_k, X_{k-1}).\end{aligned}$$

Variance of Weighted Statistic

We calculate the variance of X_α.

$$\begin{aligned}\mathrm{Var}(X_\alpha) &= \alpha^2\mathrm{Var}(X_k) + (1-\alpha)^2\mathrm{Var}(X_{k-1}) \\ &\quad + 2\alpha(1-\alpha)\mathrm{Cov}(X_k, X_{k-1}).\end{aligned}$$

Covariance calculation is involved, as it uses the joint PDF.

Covariance Calculation

Show part using joint PDF of the involved calculation.

$$\mathrm{Cov}[X_k, X_{k-1}] \quad = \quad \mathbb{E}[X_k \cdot X_{k-1}] \; - \; \mathbb{E}[X_k] \cdot \mathbb{E}[X_{k-1}].$$

$$\begin{aligned}
\mathbb{E}[X_k \cdot X_{k-1}] \quad &= \quad \mathbb{E}\Bigg[\left(m_k\left(\frac{k+1}{k}\right) - 1\right) \\
&\qquad \cdot \left(m_{k-1}\left(\frac{k+1}{k-1}\right) - 1\right)\Bigg] \\
&= \quad \frac{(k+1)^2}{k(k-1)}\mathbb{E}[m_k \cdot m_{k-1}] \; - \; \frac{k+1}{k}\mathbb{E}[m_k] \\
&\qquad - \frac{k+1}{k-1}\mathbb{E}[m_{k-1}] \; + \; \mathbb{E}[1].
\end{aligned}$$

Covariance Calculation

Use the joint PDF:

$$\begin{aligned}
\mathbb{E}[M_k \cdot M_{k-1}] &= \sum_{m_k=k}^{N} \sum_{m_{k-1}=k-1}^{m_k-1} m_k m_{k-1} \\
&\quad \mathrm{Prob}(M_{k-1} = m_{k-1}, M_k = m_k) \\
&= \sum_{m_k=k}^{N} \sum_{m_{k-1}=k-1}^{m_k-1} m_k m_{k-1} \frac{\binom{m_{k-1}-1}{k-2}}{\binom{N}{k}} \\
&\quad \text{Using identities,} \\
&= \frac{(k-1)(N+2)(N+1)}{(k+2)} - \frac{(k-1)(N+1)}{k+1}.
\end{aligned}$$

Results from Calculations

Variance of X_k

$$\mathrm{Var}(X_k) = \frac{(N-k)(N+1)}{(k)(k+2)}.$$

Variance of X_{k-1}:

$$\mathrm{Var}(X_{k-1}) = \frac{2(N-k)(N+1)}{(k+2)(k-1)}.$$

Covariance term:

$$\mathbb{E}[X_k \cdot X_{k-1}] - \mathbb{E}[X_k] \cdot \mathbb{E}[X_{k-1}] = \frac{(N+1)(N-k)}{k(k+2)}.$$

Calculating Optimal α value

Recall,

$$\begin{aligned}\mathrm{Var}(X_\alpha) &= \alpha^2\mathrm{Var}(X_k) + (1-\alpha)^2\mathrm{Var}(X_{k-1}) \\ &\quad + 2\alpha(1-\alpha)\mathrm{Cov}(X_k, X_{k-1}).\end{aligned}$$

Calculating Optimal α value

Recall,

$$\begin{aligned} \mathrm{Var}(X_\alpha) &= \alpha^2 \mathrm{Var}(X_k) + (1-\alpha)^2 \mathrm{Var}(X_{k-1}) \\ &\quad + 2\alpha(1-\alpha)\mathrm{Cov}(X_k, X_{k-1}). \end{aligned}$$

We solve for $\mathrm{Var}(X_\alpha)' = 0$ (and check the endpoints of $\alpha = 0$ or 1).

After algebra, we get the optimal α is 1.

Corresponds to using only the largest tank.

2D Square

Discrete Square Problem

Discrete Square Problem: From a square with points $(1, 1)$ to (N, N), we select k pairs without replacement.

Discrete Square Problem

Discrete Square Problem: From a square with points $(1,1)$ to (N,N), we select k pairs without replacement.

Estimate using the largest observed component.

$$\widehat{N} = \frac{2k+1}{2k}(m-1).$$

Focus on asymptotic relations as it is hard to get clean formulas.

Sample Maximum Probability & Expected Value Calculations

$$\begin{aligned}
\mathrm{PDF}_M(m) &= \mathrm{Prob}(M \le m) - \mathrm{Prob}(M \le m-1) \\
&= \frac{\binom{m^2}{k}}{\binom{N^2}{k}} - \frac{\binom{(m-1)^2}{k}}{\binom{N^2}{k}}.
\end{aligned}$$

Sample Maximum Probability & Expected Value Calculations

$$\begin{aligned}
\mathrm{PDF}_M(m) &= \mathrm{Prob}(M \leq m) - \mathrm{Prob}(M \leq m-1) \\
&= \frac{\binom{m^2}{k}}{\binom{N^2}{k}} - \frac{\binom{(m-1)^2}{k}}{\binom{N^2}{k}}.
\end{aligned}$$

Use definition of expected value, and telescope:

$$\begin{aligned}
\mathbb{E}[M] &= \sum_{m=\lceil\sqrt{k}\rceil}^{N} m \cdot \mathrm{PDF}_M(M=m) \\
&= \sum_{m=\lceil\sqrt{k}\rceil}^{N} \frac{m\binom{m^2}{k} - (m-1)\binom{(m-1)^2}{k}}{\binom{N^2}{k}} - \sum_{m=\lceil\sqrt{k}\rceil}^{N} \frac{\binom{(m-1)^2}{k}}{\binom{N^2}{k}}.
\end{aligned}$$

Euler-Maclaurin Formula

Theorem (Euler-Maclaurin formula)

For p a positive integer and a function $f(x)$ that is p times continuously differentiable on the interval $[a, b]$, we have

$$\sum_{i=a}^{b} f(i) = \int_a^b f(x)\,dx + \frac{f(a)+f(b)}{2} + \sum_{q=1}^{\lfloor \frac{p}{2} \rfloor} \frac{B_{2q}}{(2q)!}\left(f^{2q-1}(b) - f^{2q-1}(a)\right) + R_p, \text{ and}$$

$$|R_p| \leq \frac{2\zeta(p)}{(2\pi)^p} \int_m^n |f^{(p)}(x)|\,dx.$$

Bounds lemma

Lemma

For $m, L \geq 0$ and $k \geq 1$,

$$\begin{aligned} & m^{Lk} - m^{Lk-L}\left(\frac{k(k-1)}{2}\right) \\ \leq \quad & m^L(m^L-1)\cdots(m^L-(k-1)) \\ \leq \quad & m^{Lk}. \end{aligned}$$

Bounds lemma

Lemma

For $m, L \geq 0$ and $k \geq 1$,

$$\begin{aligned} & m^{Lk} - m^{Lk-L}\binom{k(k-1)}{2} \\ \leq\ & m^L(m^L-1)\cdots(m^L-(k-1)) \\ \leq\ & m^{Lk}. \end{aligned}$$

Upper bound: trivial.

Lower bound: by induction.

Then, apply Euler-Maclaurin on both bounds.

Calculations

Use only the main term to find a clean approximation (as $N \to \infty$ dominates the error terms).

Calculations

Use only the main term to find a clean approximation (as $N \to \infty$ dominates the error terms).

$$\sum_{m=\lceil\sqrt{k}\rceil}^{N-1} m^{2k} \approx \frac{(N-1)^{2k+1} - (\lceil\sqrt{k}\rceil)^{2k+1}}{2k+1}$$

Because we assumed that k is fixed,
if N is very large the other
terms are negligible.

$$\approx \frac{(N-1)^{2k+1}}{2k+1}.$$

Calculations

Now we combine all terms and get

$$\mathbb{E}[M] = N\left(\frac{2k}{2k+1}\right) + 1.$$

Invert the relationship between N and m and get

$$\widehat{N} = \frac{2k+1}{2k}(m-1).$$

Comparing Formulas

- Discrete 1D formula: $\widehat{N} = \frac{k+1}{k} \cdot m - 1.$
- Discrete 2D square formula: $\widehat{N} = \frac{2k+1}{2k}(m-1).$

Notice

$$\frac{k+1}{k} > \frac{2k+1}{2k}.$$

2D Circle

Discrete Circle Problem

Select k pairs without replacement from a circle with radius r and center (0,0).

Which statistic to study?

Discrete Circle Problem

Select k pairs without replacement from a circle with radius r and center (0,0).

Which statistic to study?

We look at $X^2 + Y^2$.

- Only integers.
- Estimate for r^2 and then take square root to estimate for r.

Gauss Circle Problem

Need to know the number of lattice points inside the circle.

Theorem (Gauss Circle Problem)

$$\text{Let } P(r) \ := \ \{\# \text{ of } (q,n) \in \mathbb{Z}^2 : q^2 + n^2 \leq r^2\}.$$

We have

$$P(r) \ = \ \pi r^2 \ + \ E(r).$$

We do not need the best known results, so we write $E(r)$ as $O(r^\delta)$. The current world record has $.5 < \delta < .63$.

Calculating the Expected Value

Calculate using the discrete CDF method. Leave them as expressions of P.

Calculating the Expected Value

Calculate using the discrete CDF method. Leave them as expressions of P.

$$
\begin{aligned}
\mathbb{E}[M] &= \sum_{m_1=0}^{r^2} m_1 \cdot \mathrm{Prob}(M = m_1) \\
&= [P(1) - P(0)] + [2P(2) - 2P(1)] + \cdots + \\
&\quad r^2[P(r^2) - r^2 P(r^2 - 1)] \\
&= r^2 P(r^2) - [P(1) + P(2) + P(r^2 - 1) \\
&= r^2 - \frac{1}{\binom{\pi r^2 + O(r^\delta)}{k}} \sum_{m_1=0}^{r^2} \binom{\pi(m_1 - 1) + O((m_1 - 1)^\delta)}{k}.
\end{aligned}
$$

Set bounds and apply E-M to calculate the second term.

Calculations

Use only the main term, as we can't invert if we use all terms.

Using similar arguments as the discrete square,

$$\sum_{m_1=0}^{r^2} \binom{\pi(m_1-1)+O(m_1^{\delta})}{k} \quad \approx \quad \pi^k \cdot \frac{(r^2-1)^{k+1}}{k+1}.$$

Now that we have a estimation for the main term, we finish our calculation.

Formula

For the expected value, we get

$$\mathbb{E}[M] \approx r^2 \cdot \frac{k}{k+1} + 1.$$

Formula

For the expected value, we get

$$\mathbb{E}[M] \quad \approx \quad r^2 \cdot \frac{k}{k+1} + 1.$$

Inverting the relationship, we get

$$\widehat{r} \;=\; \sqrt{\frac{k+1}{k}(m_1 - 1)}.$$

L-Dimensional Problem

Higher Dimensional Cases

The results and calculations for higher dimensional cases are similar to the two dimensional cases.

However, we look at different statistics:

- Generalized Discrete Square: Largest observed component.
- Generalized Discrete Circle: $X_1^2 + X_2^2 + \cdots + X_L^2$.

Results for Generalized Square

- Discrete:

$$\widehat{N} = \frac{Lk+1}{Lk}(m-1).$$

Notice $\frac{Lk+1}{Lk}$ is very very close to 1.

Scaling factor doesn't play a big role in higher dimensions.

Generalized Circle Problem

Have to know the number of lattice points inside a L-dim sphere with radius r.

$V(n)$: be the volume of a L-dim sphere.

$$V(n) \;=\; \frac{\pi^{\frac{L}{2}}}{\Gamma(\frac{L}{2}+1)} r^L :$$

Use this formula to find $P(r)$, the number of lattice points inside the L-dim sphere.

Denote the bounds with big O notation.

$$P(r) \;=\; \frac{\pi^{\frac{L}{2}}}{\Gamma(\frac{L}{2}+1)} r^L \;+\; O(r^{\delta}).$$

Results

- Discrete:

$$\widehat{r} = \sqrt{(m_1 - 1) \cdot \frac{k+1}{k}}.$$

Notice formula is independent of L.

References and Appendix

Thank you!

1. The Albert Team. (2022, March 1). *The German Tank Problem Explained: AP Statistics Review.* Albert Resources. Retrieved September 2, 2022, from https://www.albert.io/blog/german-tank-problem-explained-ap-statistics-review/
2. G. Clark, A. Gonye, and S. J. Miller, *Lessons From the German Tank Problem*. https://arxiv.org/pdf/2101.08162.pdf.
3. Constantinescu, P. (n.d.). *The Gauss Circle problem github pages. The Gauss circle problem.* Retrieved September 4, 2022, from https://petruconstantinescu.github.io/The_Gauss_circle_problem.pdf
4. Gipple, Jake (2014) *"The Volume of n-balls,"* Rose-Hulman Undergraduate Mathematics Journal: Vol. 15 : Iss. 1 , Article 14.
5. A. Hayes, *Portfolio variance*, Investopedia (2022, February 8), retrieved September 2, 2022 from https://www.investopedia.com/terms/p/portfolio-variance.asp.
6. The Investment Team, *Modern portfolio theory (MPT)*, Investopedia (2022, June 14), retrieved September 4, 2022 from https://www.investopedia.com/terms/m/modernportfoliotheory.asp.
7. M. C. Kalu, *How the allies guessed the number of German tanks using serial numbers*, warhistoryonline (2019, September 18), retrieved September 2, 2022 from https://www.warhistoryonline.com/instant-articles/the-german-tank-problem.html.
8. K. Kuter, K. *Chapter 4.1: Probability density functions (pdfs) and cumulative distribution functions (cdfs) for continuous random variables*, Statistics LibreTexts (2021, March 9), retrieved September 4, 2022 from https://tinyurl.com/2dchr2v8.
9. S. J. MILLER, *The Probability Lifesaver: All the Tools You Need to Understand Chance*, Princeton University Press (2017), https://doi.org/10.2307/j.ctvc7767n.
10. S. Sheffield, S. *18.600: Lecture 24 .1in covariance and some conditional expectation ...*, retrieved September 2, 2022 from https://math.mit.edu/~sheffield/2017600/Lecture24.pdf.
11. V. Kac and K. Yessenov, *18.704 Seminar in Algebra and Number Theory: Euler-Maclaurin Formula* (Fall 2015), retrieved September 1, 2022 from https://people.csail.mit.edu/kuat/courses/euler-maclaurin.pdf.

Simulation for 1-dim case

```
germantankcombinedestimatenolist[NumTanks_, k_, numdo_] := Module[{},
  tanks = {}; (* store list of tanks here *)
  For[i = 1, i ≤ NumTanks, i++, tanks = AppendTo[tanks, i]];
  sumX1 = 0; (* save prediction from largest tank here *)
  sumX1sq = 0; (* saves sum of squares *)
  alpha = 1; (* add formula here - function of k and NumTanks*)
  For[n = 1, n ≤ numdo, n++,
   {observedtanks = RandomSample[tanks, k]; (* uniformly at random chooses k tanks from 1 to NumTanks*)
    observedtanks = Sort[observedtanks]; (* sorts list*)
    largest = observedtanks[[-1]]; (*largest tank*)
    X1 = largest * ((k + 1.0) / k) - 1;
    sumX1 = sumX1 + X1;
    sumX1sq = sumX1sq + X1^2;
   }]; (* end of n loop*)
  Print["German Tank Problem Calculations: N = ", NumTanks, ", k = ", k, "."];
  Print["Check of Means / Variances."];
  Print["Mean(X1) = ", sumX1 1.0 / numdo, "; Var(X1) = ", (sumX1sq - sumX1^2 / numdo) / (numdo - 1.0), "."];
 ] (* end of module*)
```

```
Timing[germantankcombinedestimatenolist[400, 60, 1 000 000] ]
German Tank Problem Calculations: N = 400, k = 60.
Check of Means / Variances.
Mean(X1) = 399.989; Var(X1) = 36.781.
{5.499, Null}
```

```
Timing[germantankcombinedestimatenolist[4000, 60, 1 000 000] ]
German Tank Problem Calculations: N = 4000, k = 60.
Check of Means / Variances.
Mean(X1) = 3999.99; Var(X1) = 4245.42.
{8.69669, Null}
```

```
Timing[germantankcombinedestimatenolist[100, 2, 1 000 000] ]
German Tank Problem Calculations: N = 100, k = 2.
Check of Means / Variances.
Mean(X1) = 100.052; Var(X1) = 1234.92.
{3.5706, Null}
```

```
Timing[germantankcombinedestimatenolist[10 000, 2, 1 000 000] ]
German Tank Problem Calculations: N = 10 000, k = 2.
Check of Means / Variances.
Mean(X1) = 10 000.6; Var(X1) = 1.25123 × 10^7.
{3.89604, Null}
```

Two Dimensional Circle Code

```
discretecircle[radius_, k_, numdo_] := Module[{},
  sumestimatedradius = 0; (* save prediction from largest tank here *)
  sumestimatedradiussq = 0; (* saves sum of squares *)
iterations = 0;
  For[n = 1, n ≤ numdo, n++,
   count = 0;
R = radius;
list = {};
iterations = 0;
   {While[count < k, {iterations = iterations + 1;
      x = RandomInteger[{-R, R}]; (*Select pairs*)
      y = RandomInteger[{-R, R}]; (*Select pairs*)
      If[x^2 + y^2 ≤ R^2 && MemberQ[list, {x, y}] == False, {list = AppendTo[list, {x, y}];
        count = count + 1;}]; (*end of if statement*)}];
    findingmax = Max[Table[Total[list[[i]]^2], {i, 1, Length[list]}]];
    largestpair = Flatten[Select[list, #[[1]]^2 + #[[2]]^2 == findingmax &]]; (*Finding the largest pair*)
    beforescaled = 1.0 Sqrt[largestpair[[1]]^2 + largestpair[[2]]^2];
    estimatedradius = 1.0 Sqrt[largestpair[[1]]^2 + largestpair[[2]]^2] (2 k + 1) / (2 k);
    (*Making estimates for the radius*)
    sumestimatedradius = sumestimatedradius + estimatedradius;
    sumestimatedradiussq = sumestimatedradiussq + estimatedradius^2;
   }]; (* end of n loop*)

  Print["Mean(estimated radius) = ", sumestimatedradius 1.0 / numdo, "; Var(X1) = ",
   (sumestimatedradiussq - sumestimatedradius^2 / numdo) / (numdo - 1.0), "."];
  Print["Actual Radius = ", radius "."];
  Print["Scaled value =", 1.0 × (2 k + 1) / (2 k), "."];

 ] (* end of module*)
```

```
discretecircle[50, 5, 1000]
```

Mean(estimated radius) = 49.8962; Var(X1) = 22.5574.

Actual Radius = 50 .

Scaled value =1.1.

```
discretecircle[100, 15, 1000]
```

Mean(estimated radius) = 99.9577; Var(X1) = 10.8644.

Actual Radius = 100 .

Scaled value =1.03333.

```
discretecircle[100, 5, 1000]
```

Mean(estimated radius) = 99.6659; Var(X1) = 85.0634.

Actual Radius = 100 .

Scaled value =1.1.

Two Dimensional Discrete Square

```
discretesquare2D[NumTanks_, k_, numdo_] := Module[{},
  sumestimatednvalue = 0; (* save prediction from largest tank here *)
  sumestimatednvaluesq = 0; (* saves sum of squares *)
iterations = 0;
  For[n = 1, n ≤ numdo, n++,
   count = 0;
list = {};
iterations = 0;
   {While[count < k, {iterations = iterations + 1;
      x = RandomInteger[{0, NumTanks}]; (*Select pairs*)
      y = RandomInteger[{0, NumTanks}]; (*Select pairs*)
      If[MemberQ[list, {x, y}] == False, {list = AppendTo[list, {x, y}];
        count = count + 1;}];(*end of if statement*)}];
    maxofcomponents = Max[list];
    estimatednvalue = 1.0 (maxofcomponents - 1) (2 k + 1) / (2 k);
    sumestimatednvalue = sumestimatednvalue + estimatednvalue;
    sumestimatednvaluesq = sumestimatednvaluesq + estimatednvalue^2;
   }]; (* end of n loop*)

  Print["Mean(estimated N) = ", sumestimatednvalue 1.0 / numdo, "; Var(X1) = ",
   (sumestimatednvaluesq - sumestimatednvalue^2 / numdo) / (numdo - 1.0), "."];
  Print["Scaled value =", 1.0 × (2 k + 1) / (2 k), "."];

 ] (* end of module*)

Timing[discretesquare2D[100, 10, 100000]]

Mean(estimated N) = 99.4054; Var(X1) = 23.1241.
Scaled value =1.05.
{4.29191, Null}

Timing[discretesquare2D[100, 3, 100000]]

Mean(estimated N) = 99.2409; Var(X1) = 212.373.
Scaled value =1.16667.
{1.53073, Null}
```

Comparing 1D to 2D

N = 100, k = 20

```
In[44]:= Timing[germantankcombinedestimatenolist1D[10 000, 40, 1 000 000] ]
         Check of Means / Variances.
         Mean(X1) = 99.9936; Var(X1) = 1.51384.
Out[44]= {9.24912, Null}

In[30]:= Timing[discretesquare2D[100, 20, 1 000 000] ]
         Mean(estimated N) = 99.4301; Var(X1) = 5.97766.
         Actual Radius = 100 .
         Scaled value =1.025.
Out[30]= {87.609, Null}
```

N = 100, k = 2

```
In[48]:= Timing[germantankcombinedestimatenolist1D[10 000, 4, 1 000 000] ]
         Check of Means / Variances.
         Mean(X1) = 99.4039; Var(X1) = 123.252.
Out[48]= {4.21961, Null}

In[49]:= Timing[discretesquare2D[100, 2, 1 000 000] ]
         Mean(estimated N) = 99.1212; Var(X1) = 425.357.
         Scaled value =1.25.
Out[49]= {11.5519, Null}
```

N = 200, k = 30

```
In[•]:= Timing[germantankcombinedestimatenolist1D[40 000, 60, 1 000 000] ]
        Check of Means / Variances.
        Mean(X1) = 199.992; Var(X1) = 2.73095.
Out[•]= {11.7437, Null}

In[•]:= Timing[discretesquare2D[200, 30, 1 000 000] ]
        Mean(estimated N) = 199.452; Var(X1) = 10.7529.
        Actual Radius = 200 .
        Scaled value =1.01667.
Out[•]= {136.32, Null}
```

N = 30, k = 5

```
In[•]:= Timing[germantankcombinedestimatenolist1D[900, 10, 1 000 000] ]
        Check of Means / Variances.
        Mean(X1) = 29.9673; Var(X1) = 2.01427.
Out[•]= {4.71964, Null}

In[•]:= Timing[discretesquare2D[30, 5, 1 000 000] ]
        Mean(estimated N) = 29.3293; Var(X1) = 7.86311.
        Actual Radius = 30 .
        Scaled value =1.1.
Out[•]= {23.2372, Null}
```