Exercise Walkthrough: Server Overload

Justin Lanfermann

25. June 2025

Abstract

This document provides a detailed, step-by-step walkthrough for an exercise on discrete probability theory. We will calculate properties of sums of random variables, apply fundamental probability inequalities, and use the Poisson distribution for modeling and exact calculations. Each step is explained with reference to core concepts from the lecture script.

1 Problem Statement

A web server receives requests from two types of users: Type A and Type B. Let N_A be the number of requests from Type A users in an hour, with $\mathbb{E}[N_A] = 50$ and $\mathrm{var}[N_A] = 50$. Similarly, N_B is the number of requests from Type B users in an hour, with $\mathbb{E}[N_B] = 100$ and $\mathrm{var}[N_B] = 100$. Assume N_A and N_B are independent. The server has a capacity to handle $C = 200$ requests per hour. If the total number of requests $N_T = N_A + N_B$ exceeds C, the server is overloaded and breaks down.

(i) Calculate the expected total number of requests per hour, $\mathbb{E}[N_T]$, and the variance of the total number of requests per hour, $\mathrm{var}[N_T]$.

(ii) Using Chebyshev's inequality, provide an upper bound on the probability that the server is overloaded, i.e., on $P(N_T > C)$.

(iii) Assuming N_T is non-negative (quite plausibly), use Markov's inequality and $\mathbb{E}[N_T]$ to provide another upper bound on $P(N_T > C)$. Compare it to the bound in (ii) and explain the difference.

(iv) What would be an adequate distribution to model N_A, N_B and why?

(v) One can show that for $X \sim \mathrm{Poi}(\lambda_1)$, $Y \sim \mathrm{Poi}(\lambda_2)$ and $X \perp Y$, we have that $X + Y \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$. Using this result, calculate the probability of server overload $P(N_T > C)$ under this model. (You can use a computer for this.) Compare this value with the bounds obtained in (ii) and (iii). Discuss the differences and the utility of the bounds versus an exact calculation for a specific distribution.

2 Step-by-Step Solution

2.1 (i) Expectation and Variance of Total Requests

Overview: We need to find the mean and variance of the sum of two random variables, $N_T = N_A + N_B$. We will use the property of linearity of expectation and the property for the variance of a sum of *independent* random variables.

Step-by-step Derivation:

1. **Expected Value $\mathbb{E}[N_T]$:** The expectation of a sum of random variables is always the sum of their individual expectations. This is known as the linearity of expectation. [1]

   - **Concept Used:** Linearity of Expectation (Proposition 2.4 (i) in the script).
   - **Calculation**:

   $$\mathbb{E}[N_T] = \mathbb{E}[N_A + N_B]$$
   $$= \mathbb{E}[N_A] + \mathbb{E}[N_B]$$
   $$= 50 + 100 = 150.$$

   So, the expected total number of requests per hour is 150.

2. **Variance $\mathbf{var}[N_T]$:** The variance of a sum of *independent* random variables is the sum of their variances. The independence of N_A and N_B is crucial here. [2]

   - **Concept Used:** Variance of a sum of independent variables (Proposition 2.8 (iv) in the script).
   - **Reasoning:** The problem states that N_A and N_B are independent. This allows us to simply add their variances. If they were not independent, we would need to account for their covariance: $\text{var}[N_A + N_B] = \text{var}[N_A] + \text{var}[N_B] + 2\text{cov}[N_A, N_B]$.
   - **Calculation**:

   $$\text{var}[N_T] = \text{var}[N_A + N_B]$$
   $$= \text{var}[N_A] + \text{var}[N_B] \quad \text{(due to independence)}$$
   $$= 50 + 100 = 150.$$

   The variance of the total number of requests is also 150.

2.2 (ii) Bound using Chebyshev's Inequality

Overview: We will use Chebyshev's inequality to find an upper bound for the probability of server overload, $P(N_T > 200)$. This inequality provides a bound on the probability that a random variable deviates from its mean by a certain amount, and only requires the mean and variance. [3]

Step-by-step Derivation:

1. **Formulate the Event:** The server overloads if $N_T > 200$. Since the number of requests must be an integer, this is equivalent to $N_T \geq 201$. We want to bound this probability.

2. **Relate to Chebyshev's Form:** Chebyshev's inequality is stated as $P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$. We need to express our event, $N_T \geq 201$, in this form.

- We know $\mu = \mathbb{E}[N_T] = 150$.
- The event $N_T \geq 201$ is equivalent to $N_T - 150 \geq 51$.
- The event $\{N_T - 150 \geq 51\}$ is a subset of the event $\{|N_T - 150| \geq 51\}$. Therefore, $P(N_T - 150 \geq 51) \leq P(|N_T - 150| \geq 51)$. This allows us to apply the inequality.

3. **Apply the Inequality:**

- **Concept Used:** Chebyshev's Inequality (Theorem 2.40 in the script).
- **Calculation:** We set $\epsilon = 51$.

$$
\begin{aligned}
P(N_T > 200) &= P(N_T \geq 201) \\
&= P(N_T - 150 \geq 51) \\
&\leq P(|N_T - 150| \geq 51) \\
&\leq \frac{\text{var}[N_T]}{\epsilon^2} \\
&= \frac{150}{51^2} = \frac{150}{2601} \approx 0.05767.
\end{aligned}
$$

Conclusion: The probability of a server overload is at most about 5.77%.

2.3 (iii) Bound using Markov's Inequality and Comparison

Overview: We use Markov's inequality to find another upper bound for $P(N_T > 200)$. This inequality is more general than Chebyshev's but uses less information (only the mean and the fact that the variable is non-negative), often resulting in a looser bound. [4]

Step-by-step Derivation:

1. **Verify Conditions:** Markov's inequality, $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$, applies to non-negative random variables. The total number of requests N_T is indeed non-negative.

2. **Apply the Inequality:**

- **Concept Used:** Markov's Inequality (Theorem 2.38 in the script).
- **Calculation:** We want to bound $P(N_T > 200)$. In the form $P(X \geq a)$, this is $P(N_T \geq 201)$ or we can directly use $a = 200$ for $P(N_T > 200)$. The standard form uses $\geq$, so let's stick to $P(N_T \geq 201)$.

$$
P(N_T > 200) = P(N_T \geq 201) \leq \frac{\mathbb{E}[N_T]}{201} = \frac{150}{201} \approx 0.746
$$

(Note: Applying it with $a = 200$ gives $P(N_T > 200) \leq \frac{150}{200} = 0.75$, which is also a valid, slightly looser bound).

Comparison and Explanation:

- **Chebyshev's Bound:** $\approx 5.77\%$

- **Markov's Bound:** $\approx 75\%$

The Chebyshev bound is significantly *tighter* (smaller, and therefore more informative) than the Markov bound. This is expected. Markov's inequality only uses the mean ($\mathbb{E}[N_T]$) and the fact that $N_T \geq 0$. Chebyshev's inequality uses both the mean *and* the variance ($\text{var}[N_T]$). The variance provides crucial information about the spread or concentration of the distribution around its mean. Since we used more information, we got a better (tighter) bound.

2.4 (iv) Choosing a Distributional Model

Overview: We need to propose a suitable probability distribution for N_A and N_B based on the problem description and common modeling practices in probability theory.

Justification: The **Poisson distribution** is an excellent model for N_A and N_B. [5]

- **Physical Intuition:** A Poisson distribution models the number of events (requests) occurring in a fixed interval of time or space, assuming these events happen with a known constant mean rate and independently of the time since the last event. This is a standard and often effective model for phenomena like website requests, customer arrivals, or radioactive decays.

- **Mathematical Justification:** A key property of a Poisson-distributed random variable $X \sim \text{Poi}(\lambda)$ is that its expectation and variance are both equal to λ.

  - For Type A users: $\mathbb{E}[N_A] = 50$ and $\text{var}[N_A] = 50$. This perfectly matches a Poisson distribution with $\lambda_A = 50$.
  - For Type B users: $\mathbb{E}[N_B] = 100$ and $\text{var}[N_B] = 100$. This perfectly matches a Poisson distribution with $\lambda_B = 100$.

  This alignment of mean and variance given in the problem statement is a very strong hint that the Poisson model is intended.

Conclusion: We model $N_A \sim \text{Poi}(50)$ and $N_B \sim \text{Poi}(100)$.

2.5 (v) Exact Calculation and Final Discussion

Overview: Using the Poisson model from part (iv) and the given property that the sum of independent Poisson variables is also Poisson, we will calculate the exact probability of server overload. We then compare this exact value to the bounds from (ii) and (iii).

Step-by-step Derivation:

1. **Find the Distribution of N_T:**

   - **Concept Used:** Sum of independent Poisson random variables (provided in the problem, also seen in Example 2.50 (i) of the script).
   - **Derivation:** Since $N_A \sim \text{Poi}(50)$, $N_B \sim \text{Poi}(100)$, and they are independent, their sum is:
     $$N_T = N_A + N_B \sim \text{Poi}(50 + 100) = \text{Poi}(150)$$

2. **Calculate the Exact Probability $P(N_T > 200)$:**

   - The probability mass function (pmf) for $N_T \sim \text{Poi}(150)$ is $P(N_T = k) = \frac{e^{-150}150^k}{k!}$.
   - We want to calculate $P(N_T > 200)$, which is $\sum_{k=201}^{\infty} P(N_T = k)$.
   - It is much easier to calculate this using the cumulative distribution function (CDF):

     $$P(N_T > 200) = 1 - P(N_T \leq 200) = 1 - \sum_{k=0}^{200} \frac{e^{-150}150^k}{k!}$$

   - **Computation:** This sum is tedious to compute by hand. Using a computer (e.g., Python's 'scipy.stats.poisson'), we get:

```
from scipy.stats import poisson
# P(NT > 200) = 1 - P(NT <= 200)
# The parameters are k (the value) and lambda (the mean)
prob = 1 - poisson.cdf(200, 150)
# prob is approx 0.000042
```

So, $P(N_T > 200) \approx 0.000042$, or 0.0042%.

Discussion and Comparison:

- **Markov's Bound:** $P(\text{overload}) \leq 75\%$

- **Chebyshev's Bound:** $P(\text{overload}) \leq 5.77\%$

- **Poisson Model Exact Probability:** $P(\text{overload}) \approx 0.0042\%$

This comparison is very instructive. The bounds from Markov's and Chebyshev's inequalities are extremely loose compared to the exact probability calculated under the Poisson assumption. This highlights a fundamental trade-off in probabilistic modeling:

- **Inequalities (Bounds):** They are powerful because they are *distribution-free*. They require very few assumptions (mean for Markov, mean and variance for Chebyshev) and the bounds they provide are always correct, regardless of the true underlying distribution. Their weakness is that they can be very pessimistic (loose), especially for distributions that are well-behaved and concentrated around their mean (like the Poisson distribution in this high-lambda regime).

- **Specific Models (Exact Calculation):** By assuming a specific distribution (Poisson), we can get a precise, quantitative answer. This answer, however, is only as reliable as our modeling assumption. If the real-world process of receiving requests deviates significantly from a Poisson process (e.g., requests come in bursts, violating independence), then our "exact" probability could be misleadingly optimistic.

In summary, the inequalities provide a robust, worst-case sanity check, while specific distributional models offer precision at the cost of being potentially wrong if the assumptions do not hold.

3 In-depth Explanations

[1] Linearity of Expectation

This is a fundamental property of the expectation operator. For any two random variables X and Y (defined on the same probability space) and any two constants $a, b \in \mathbb{R}$, it states that:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Crucially, this property holds regardless of whether X and Y are independent or not. It is one of the most powerful tools in probability theory because of its generality. This is stated in **Proposition 2.4 (i)** of your script.

[2] Variance of a Sum of Independent Variables

For any two *independent* random variables X and Y, the variance of their sum is the sum of their variances:

$$\text{var}[X + Y] = \text{var}[X] + \text{var}[Y]$$

This is a special case of the more general formula for any two variables: $\text{var}[X + Y] = \text{var}[X] + \text{var}[Y] + 2\text{cov}[X, Y]$. When X and Y are independent, their covariance is zero, and the formula simplifies. This property is found in **Proposition 2.8 (iv)**.

[3] Chebyshev's Inequality

This inequality provides an upper bound on the probability that a random variable X with finite mean μ and finite non-zero variance σ^2 will be at a distance of $\epsilon > 0$ or more from its mean. The formula is:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

It is powerful because it connects the "spread" of a distribution (its variance) to the probability of observing outcomes far from the center, without needing to know anything else about the shape of the distribution. It is stated in **Theorem 2.40**.

[4] Markov's Inequality

This is an even more fundamental inequality. For any non-negative random variable X (i.e., $P(X \geq 0) = 1$) with finite mean $\mathbb{E}[X]$, and for any constant $a > 0$, it states:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

It provides a (often loose) upper bound on the probability of a large value, using only the mean. Its proof is surprisingly simple, but it is the foundation upon which other inequalities, including Chebyshev's, are built. It is stated in **Theorem 2.38**.

[5] The Poisson Distribution

The Poisson distribution, denoted $\text{Poi}(\lambda)$, describes the probability of a given number of events occurring in a fixed interval of time or space. Its parameter λ is a positive real number, representing the average rate of occurrence. The probability mass function (pmf) is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

A defining characteristic, as noted in **Lemma 2.26 (vii)**, is that its expected value and its variance are both equal to its parameter λ:

$$\mathbb{E}[X] = \lambda \quad \text{and} \quad \text{var}[X] = \lambda$$

$$\mathbb{E}[X] = \lambda \quad \text{and} \quad \text{var}[X] = \lambda$$