EC304: Probability Theory and Stochastic Process

Module 4: Joint Distributions and Statistical Bounds

1 Joint Distributions of Random Variables

When we consider two random variables X and Y together, their joint distribution shows the probability that both take certain values simultaneously.

$$P(X = x_i, Y = y_j)$$

Example 1 (Two Dice): Let X = outcome of die 1, Y = outcome of die 2. Each pair (x, y), where $x, y \in \{1, 2, 3, 4, 5, 6\}$, has probability:

$$P(X = x, Y = y) = \frac{1}{36}$$

The **marginal distributions** are:

$$P(X = x) = \sum_y P(X = x, Y = y) \quad \text{and} \quad P(Y = y) = \sum_x P(X = x, Y = y)$$

Practical Analogy: If X = temperature (°C) and Y = humidity (%) measured by sensors at the same time, then the joint distribution shows how often both temperature and humidity combinations occur.

—

2 Conditional Distributions and Densities

Conditional probability tells us the likelihood of one variable given another.

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

Example (Weather Sensor): Let X = rain indicator (1 if raining, 0 if not), and Y = humidity level. If humidity is high, the probability of rain increases:

$$P(X = 1 \mid Y = \text{high}) > P(X = 1 \mid Y = \text{low})$$

This concept is used in **machine learning**, **Bayesian networks**, and **weather forecasting**.

—

3 Inequalities and Bounds

Sometimes, we don't know the full probability distribution of a random variable or it's too complex to calculate exactly.

In such cases, these inequalities help us quickly estimate how likely it is for a random variable to take large or extreme values, using only basic information like mean or variance.

3.1 Markov's Inequality

For any non-negative random variable X and for any $a > 0$, the probability that X exceeds a is bounded by the ratio of its mean to a:

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

This gives an **upper bound** on the probability that X takes a large value, even when the full probability distribution is unknown.

3.1.1 Intuitive Meaning

Markov's inequality provides a simple way to understand how extreme events are limited by the average value. If the average value $E[X]$ is small, the probability that X is very large must also be small.

3.1.2 Examples

Example 1: Battery Life of a Device
Let X = battery life (hours), and the average battery life $E[X] = 5$. We want to know how likely it is that the battery lasts more than 10 hours:

$$P(X \geq 10) \leq \frac{5}{10} = 0.5$$

Interpretation: At most 50% of batteries can last more than 10 hours.

Example 2: Internet Download Time
Let X = download time (minutes), $E[X] = 4$. Probability that download time exceeds 12 minutes:

$$P(X \geq 12) \leq \frac{4}{12} = \frac{1}{3}$$

Interpretation: There is at most a 33% chance that the download will take longer than 12 minutes.

Example 3: Packet Delay in a Network
Suppose packet delay X has mean $E[X] = 50$ ms. Find an upper bound for $P(X \geq 200)$:

$$P(X \geq 200) \leq \frac{50}{200} = 0.25$$

Interpretation: No more than 25% of packets can have delay greater than 200 ms. Useful in *Quality of Service (QoS)* analysis.

Example 4: Factory Production Time

Let X = time (minutes) to manufacture one unit, $E[X] = 10$. Bound the probability that it takes longer than 40 minutes:

$$P(X \geq 40) \leq \frac{10}{40} = 0.25$$

Interpretation: At most 25% of products take longer than 40 minutes. Helps in production planning.

Example 5: Sensor Measurement Noise

Let X represent the absolute error of a temperature sensor, $E[X] = 0.5°C$. Find the upper bound for error exceeding $2°C$:

$$P(X \geq 2) \leq \frac{0.5}{2} = 0.25$$

Interpretation: At most 25% of readings can have error greater than $2°C$. Used in sensor calibration and reliability testing.

Example 6: CPU Execution Time

If average execution time $E[X] = 2$ seconds, probability that it exceeds 6 seconds:

$$P(X \geq 6) \leq \frac{2}{6} = 0.333$$

Interpretation: There is at most a 33% chance that a process takes longer than 6 seconds.

Example 7: Machine Uptime

Suppose machine uptime $E[X] = 100$ hours. Bound the probability that uptime exceeds 300 hours:

$$P(X \geq 300) \leq \frac{100}{300} = \frac{1}{3}$$

Interpretation: Only up to one-third of machines can be expected to run beyond 300 hours without failure.

3.2 Chebyshev's Inequality

For any random variable X with mean μ and variance σ^2, and for any $k > 0$:

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

This gives a bound on how much X deviates from its mean. It tells us that the probability of large deviations from the mean becomes smaller as k increases.

Example 1: Sensor Temperature Readings

Let X be the temperature (in °C) recorded by a sensor, with mean $\mu = 25$ and standard deviation $\sigma = 2$. Find $P(|X - 25| \geq 6)$.

$$P(|X-25| \geq 6) \leq \frac{1}{k^2} = \frac{1}{(6/2)^2} = \frac{1}{9} \approx 0.111$$

Interpretation: At most 11.1% of temperature readings differ from the mean by more than 6°C.

Example 2: Internet Download Speed

Let X be the download speed (Mbps) with $\mu = 50$, $\sigma = 5$. Find the probability that the speed differs from the mean by more than 15 Mbps.

$$P(|X-50| \geq 15) \leq \frac{1}{(15/5)^2} = \frac{1}{9} \approx 0.111$$

Interpretation: Only up to 11.1% of speed readings deviate from the mean by more than 15 Mbps.

Example 3: Student Exam Scores

Let X be students' marks with mean $\mu = 70$, $\sigma = 10$. We want $P(|X - 70| \geq 20)$.

$$P(|X-70| \geq 20) \leq \frac{1}{(20/10)^2} = \frac{1}{4} = 0.25$$

Interpretation: At most 25% of students score outside the range 50–90.

Example 4: Battery Capacity

Let X = battery capacity (mAh) with $\mu = 3000$, $\sigma = 300$. Find $P(|X - 3000| \geq 900)$.

$$P(|X-3000| \geq 900) \leq \frac{1}{(900/300)^2} = \frac{1}{9} \approx 0.111$$

Interpretation: Only about 11% of batteries deviate by more than 900 mAh from the average capacity.

Example 5: Machine Processing Time

Let X = time (seconds) to process a task, $\mu = 8$, $\sigma = 2$. Find $P(|X - 8| \geq 4)$.

$$P(|X - 8| \geq 4) \leq \frac{1}{(4/2)^2} = \frac{1}{4} = 0.25$$

Interpretation: At most 25% of tasks take less than 4 seconds or more than 12 seconds to complete.

3.3 Chernoff Bound

The **Chernoff Bound** provides a very tight upper bound on the probability that a random variable deviates significantly from its expected value. It is especially useful for sums of independent random variables, such as coin tosses, sensor readings, or packet arrivals in a network.

It gives an *exponentially decreasing bound* on tail probabilities, meaning that large deviations are extremely unlikely.

3.3.1 Mathematical Form

Let $X_1, X_2, ..., X_n$ be independent random variables taking values in $[0, 1]$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$.

Then, for any $\delta > 0$:

$$P(X \geq (1 + \delta)\mu) \leq \left(\frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right)^{\mu}$$

This bound decreases exponentially with μ and δ.

Example 1: Coin Toss

Suppose we flip a fair coin 100 times. Each flip has probability $p = 0.5$ of getting heads. Hence, the expected number of heads is:

$$\mu = np = 100 \times 0.5 = 50$$

We want the probability of getting 70 or more heads.

$$(1 + \delta)\mu = 70 \Rightarrow 1 + \delta = \frac{70}{50} = 1.4 \Rightarrow \delta = 0.4$$

Now applying Chernoff bound:

$$P(X \geq 70) \leq \left(\frac{e^{0.4}}{1.4^{1.4}} \right)^{50} \approx 0.0003$$

Interpretation: There is less than 0.03% chance of getting 70 or more heads — showing that large deviations from the mean are highly unlikely.

Example 2: Network Packet Arrivals

Suppose a router receives $n = 200$ packets per second on average. So, $\mu = 200$. We want the probability that it receives 300 or more packets in a second.

$$(1+\delta)\mu = 300 \Rightarrow \delta = \frac{300-200}{200} = 0.5$$

Applying the bound:

$$P(X \geq 300) \leq \left(\frac{e^{0.5}}{1.5^{1.5}}\right)^{200} \approx 1.4 \times 10^{-14}$$

Interpretation: The probability that packet traffic exceeds 300 packets in a second is almost zero. Hence, Chernoff bounds are very useful in **network traffic analysis and system design**.

Example 3: Sensor Noise Detection

Assume a temperature sensor gives binary readings (1 if noise spike detected, 0 otherwise) over $n = 1000$ readings. Suppose the probability of a spike is $p = 0.01$, so $\mu = np = 10$.

We want the probability that more than 20 spikes occur:

$$(1+\delta)\mu = 20 \Rightarrow \delta = 1$$

Hence,

$$P(X \geq 20) \leq \left(\frac{e^1}{2^2}\right)^{10} = \left(\frac{2.718}{4}\right)^{10} \approx 0.06$$

Interpretation: There is only about a 6% chance of seeing more than twice the expected number of noise spikes — showing that large bursts of noise are rare.