Lecture Notes:
Probabilistic Inequalities, Convexity, and Sampling Methods

Lecture Notes for Undergraduate Probability

April 22, 2025

Announcements:

- Good evening, everyone. Welcome! Let's give everyone a moment to settle in.

- We plan to take a short break approximately halfway through today's lecture.

- **Homework Clarification:** For the current assignment, Question 3 asks you to implement some of the sampling methods we'll discuss today. You'll generate random samples and use tools like histograms to visually compare the generated data's distribution to the theoretical probability density or mass function. This exercise emphasizes the practical simulation aspect of these techniques.

- **Past Exam Questions:** Some of you have asked about using past exam questions for practice. If you have access to exams from previous iterations of this course (perhaps those taught by Prof. Ben Yakir or others), you are welcome to use them for study. In about two weeks, once we have covered more foundational material, feel free to share specific questions with me, and I can help identify which ones are relevant to our current syllabus. However, please understand that I cannot provide full solutions to past exams. The goal is for you to practice applying the concepts yourself.

---

1 Probabilistic Inequalities: Bounding Probabilities

Often in probability and statistics, we need to understand certain aspects of a random variable's behavior (like how likely it is to be far from its average) even when we don't know its exact probability distribution. Probabilistic inequalities are powerful tools that allow us to establish bounds on probabilities using only limited information, such as the random variable's mean or variance.

1.1 Markov's Inequality: A Bound Using the Mean

Our first inequality, Markov's inequality, provides a fundamental (though sometimes loose) upper bound on the probability that a *non-negative* random variable takes on a value significantly

larger than its expectation. It formalizes the intuition that if the average value is small, extremely large values must be relatively unlikely.

Theorem 1.1 (Markov's Inequality). *Let X be a random variable such that $X \geq 0$ (almost surely). Let $\mathbb{E}[X]$ denote its expectation. Then for any constant $a > 0$, we have:*

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

Proof. The proof relies on the definition of expectation and the non-negativity of X. Let's consider the case where X is a continuous random variable with probability density function (PDF) $f(x)$. The proof for a discrete variable is analogous using sums instead of integrals.

The expectation is defined as $\mathbb{E}[X] = \int_0^\infty x f(x) dx$ (the integral starts at 0 since $X \geq 0$). We can split this integral at the point a:

$$\mathbb{E}[X] = \int_0^a x f(x) dx + \int_a^\infty x f(x) dx$$

Since $x \geq 0$ and $f(x) \geq 0$, both integrals are non-negative. Therefore, we can drop the first integral to get an inequality:

$$\mathbb{E}[X] \geq \int_a^\infty x f(x) dx$$

Now, within the range of integration ($x \geq a$), we know that x is always greater than or equal to the constant a. So, we can replace x with a inside the integral, which further reduces (or keeps equal) the value of the integral:

$$\int_a^\infty x f(x) dx \geq \int_a^\infty a f(x) dx = a \int_a^\infty f(x) dx$$

The remaining integral, $\int_a^\infty f(x) dx$, is precisely the definition of the probability $\mathbb{P}(X \geq a)$. Combining the inequalities, we have:

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

Since $a > 0$, we can divide by a without changing the direction of the inequality, yielding the desired result:

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

$\square$

Remark 1.2. Markov's inequality is very general, requiring only non-negativity and a finite mean. However, because it uses so little information, the bound it provides might not be very tight in practice.

1.2 A Corollary to Markov: Tail Decay and Higher Moments

Markov's inequality can be cleverly extended to relate the existence of higher moments of X (like $\mathbb{E}[X^2]$, $\mathbb{E}[X^3]$, etc.) to how quickly the "tail probability" $\mathbb{P}(X \geq x)$ goes to zero as x becomes very large.

Corollary 1.3 (Tail Probability Bound via Moments). *Let X be a non-negative random variable ($X \geq 0$). If $\mathbb{E}[|X|^p] = \mathbb{E}[X^p] < \infty$ for some constant $p > 0$, then for any $x > 0$:*

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X^p]}{x^p}$$

Proof. This result follows directly from applying Markov's inequality itself. Consider the random variable $Y = X^p$. Since $X \geq 0$ and $p > 0$, Y is also a non-negative random variable, $Y \geq 0$. Its expectation is $\mathbb{E}[Y] = \mathbb{E}[X^p]$, which is finite by assumption. Now, observe that because $x > 0$ and the function $t \mapsto t^p$ is strictly increasing for $t \geq 0$ when $p > 0$, the event $\{X \geq x\}$ is exactly the same as the event $\{X^p \geq x^p\}$. That is, $\{X \geq x\} \iff \{Y \geq x^p\}$. We can apply Markov's inequality (Theorem 1.1) to the non-negative random variable $Y = X^p$, using the value $a = x^p$ (which is positive since $x > 0$):

$$\mathbb{P}(Y \geq x^p) \leq \frac{\mathbb{E}[Y]}{x^p}$$

Substituting back $Y = X^p$ and using the equivalence of the events:

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X^p]}{x^p}$$

$\square$

Remark 1.4 (Interpretation: Rate of Decay). This corollary is quite insightful. It tells us that if X has a finite p-th moment ($E[X^p] < \infty$), then the probability that X exceeds a large value x must decrease at least as fast as x^{-p} (since $E[X^p]$ is just a constant). The higher the moments that exist (i.e., the larger p can be), the faster the tail probability must decay towards zero. This connects the integrability (existence of moments) of a random variable to the behavior of its tails. The proof technique essentially mirrors the original Markov proof, starting from $E[X^p]$ instead of $E[X]$.

1.3 Chebyshev's Inequality: A Bound Using Mean and Variance

Markov's inequality ignores information about the spread or variability of X. Chebyshev's inequality incorporates the variance ($\sigma^2 = \text{Var}(X)$) along with the mean ($\mu = \mathbb{E}[X]$) to provide a bound on the probability that a random variable deviates from its mean by a certain amount. It usually gives a tighter bound than Markov's inequality when applied appropriately.

Theorem 1.5 (Chebyshev's Inequality). *Let X be a random variable with finite mean $\mu = \mathbb{E}[X]$ and finite variance $\sigma^2 = \text{Var}(X)$. Then for any constant $k > 0$:*

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\text{Var}(X)}{k^2} = \frac{\sigma^2}{k^2}$$

Proof. The proof elegantly uses Markov's inequality. Consider the random variable $Y = (X - \mu)^2$. This variable measures the squared deviation of X from its mean. Crucially, Y is always non-negative, $Y \geq 0$. Its expectation is $\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2]$, which is the definition of the variance of X, so $\mathbb{E}[Y] = \text{Var}(X) = \sigma^2$. Now, consider the event $\{|X - \mu| \geq k\}$. Since $k > 0$ and the squaring function is non-decreasing for non-negative values, this event is equivalent to the event $\{(X - \mu)^2 \geq k^2\}$. That is, $\{|X - \mu| \geq k\} \iff \{Y \geq k^2\}$. We can apply Markov's inequality

(Theorem 1.1) to the non-negative variable $Y = (X - \mu)^2$, using the value $a = k^2$ (which is positive since $k > 0$):

$$\mathbb{P}(Y \geq k^2) \leq \frac{\mathbb{E}[Y]}{k^2}$$

Substituting the equivalent event and the expectation of Y:

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\text{Var}(X)}{k^2}$$

This completes the proof. □

Remark 1.6 (Standard Deviation Form). A common way to state Chebyshev's inequality is in terms of standard deviations. By setting $k = c\sigma$ where $c > 0$ is the number of standard deviations, the inequality becomes:

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

For example, the probability of being 2 or more standard deviations away from the mean is at most $1/2^2 = 1/4$, regardless of the specific distribution (as long as mean and variance are finite).

Remark 1.7 (Generality vs. Tightness). Like Markov's inequality, Chebyshev's strength lies in its generality. It applies to *any* distribution with a finite mean and variance. However, for specific known distributions (like the normal distribution, where we know $\approx 5\%$ lies beyond 2σ), the bound $1/c^2$ might be quite conservative.

1.4 Corollary to Chebyshev: Concentration of the Sample Mean

Chebyshev's inequality is a key ingredient in proving fundamental results like the Law of Large Numbers. This corollary shows how the sample mean concentrates around the true mean as the sample size grows.

Corollary 1.8 (Weak Law of Large Numbers Variant). *Let* $X_1, X_2, \ldots, X_n$ *be **uncorrelated** random variables, all having the same finite mean* $\mu = \mathbb{E}[X_i]$ *and the same finite variance* $\sigma^2 = \text{Var}(X_i)$. *Let* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ *be the sample mean. Then, for any constant* $\epsilon > 0$:

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

Proof. We will apply Chebyshev's inequality (Theorem 1.5) to the random variable $\bar{X}_n$. To do this, we first need its mean and variance.

Using linearity of expectation:

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \mu$$

So, the sample mean $\bar{X}_n$ is an unbiased estimator of the true mean μ.

Now for the variance. Recall that $\text{Var}(cY) = c^2\text{Var}(Y)$.

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n}X_i\right)$$

Because the variables $X_1,\ldots,X_n$ are uncorrelated, the variance of their sum is the sum of their variances (the covariance terms are zero):

$$\text{Var}\left(\sum_{i=1}^{n}X_i\right) = \sum_{i=1}^{n}\text{Var}(X_i) = \sum_{i=1}^{n}\sigma^2 = n\sigma^2$$

Substituting this back:

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

The variance of the sample mean decreases as n increases.

Now we apply Chebyshev's inequality to $\bar{X}_n$, noting that its mean is μ and its variance is σ^2/n. For any $\epsilon > 0$:

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2}$$

Substituting the mean and variance we just calculated:

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

$\square$

Remark 1.9 (Significance). This result is a cornerstone of statistical inference. It shows that as the sample size n gets large, the probability that the sample mean $\bar{X}_n$ deviates from the true population mean μ by more than any fixed positive amount ϵ approaches zero (since $\frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$). The probability bound decreases proportionally to $1/n$. This provides rigorous justification for using the sample mean to estimate the population mean. Note that this only requires the variables to be uncorrelated, which is a weaker condition than full independence.

2 Expectation and Convex Functions

Convexity is a geometric property of functions that has profound implications when combined with the concept of expectation. The key result is Jensen's inequality.

2.1 Review of Convex Functions

Let's formally define convexity and related concepts.

Definition 2.1 (Convex and Concave Functions). Let $I \subseteq \mathbb{R}$ be an interval. A function $g : I \to \mathbb{R}$ is called **convex** on I if for every $x, y \in I$ and every $\lambda \in [0,1]$,

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$$

Geometrically, the function's graph lies below the chord connecting any two points on the graph.

A function g is **concave** on I if the inequality is reversed, i.e., for every $x, y \in I$ and $\lambda \in [0, 1]$,

$$g(\lambda x + (1 - \lambda)y) \geq \lambda g(x) + (1 - \lambda)g(y)$$

Equivalently, g is concave if $-g$ is convex.

Remark 2.2 (Checking Convexity). For twice-differentiable functions g, convexity can be checked using the second derivative:

- If $g''(x) \geq 0$ for all x in the interior of I, then g is convex on I.

- If $g''(x) \leq 0$ for all x in the interior of I, then g is concave on I.

Examples: $g(x) = x^2 \implies g''(x) = 2 \geq 0$ (convex). $g(x) = e^x \implies g''(x) = e^x \geq 0$ (convex). $g(x) = \ln(x) \implies g''(x) = -1/x^2 \leq 0$ for $x > 0$ (concave). $g(x) = \sqrt{x} \implies g''(x) = -1/4x^{-3/2} \leq 0$ for $x > 0$ (concave).

Lemma 2.3 (Supporting Lines for Convex Functions). *A function $g : I \to \mathbb{R}$ is convex if and only if for each x_0 in the interior of I, there exists a real number $v(x_0)$ (representing the slope of a supporting line at x_0) such that*

$$g(x) \geq g(x_0) + v(x_0)(x - x_0) \quad \text{for all } x \in I.$$

If g is differentiable at x_0, then $v(x_0) = g'(x_0)$. The line $y = g(x_0) + v(x_0)(x - x_0)$ is tangent to or below the graph of g.

(The proof details were sketched in the lecture; we will use this property.)

2.2   Jensen's Inequality: Connecting Convexity and Expectation

Jensen's inequality provides a fundamental relationship between the expectation of a convex transformation of a random variable and the transformation of its expectation.

Theorem 2.4 (Jensen's Inequality). *Let X be a random variable whose values lie in an interval I. Let $\mu = \mathbb{E}[X]$, and assume $\mu \in I$. Let $g : I \to \mathbb{R}$ be a **convex** function. If both $\mathbb{E}[X]$ and $\mathbb{E}[g(X)]$ exist, then*

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

*If g is **concave**, the inequality is reversed: $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$.*

Proof. We utilize the supporting line property of convex functions (Lemma 2.3). Since g is convex and $\mu = \mathbb{E}[X]$ is in its domain I, there exists a slope $v(\mu)$ such that for all possible values x that X can take:

$$g(x) \geq g(\mu) + v(\mu)(x - \mu)$$

Since this inequality holds for every possible value x of the random variable X, it must also hold *for the random variable X itself*:

$$g(X) \geq g(\mu) + v(\mu)(X - \mu)$$

Now, we take the expectation of both sides. Since the inequality holds for the random variables, it holds for their expectations (properties of expectation):

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(\mu) + \nu(\mu)(X - \mu)]$$

Using the linearity of expectation on the right-hand side:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(\mu)] + \mathbb{E}[\nu(\mu)(X - \mu)]$$

Note that $\mu = \mathbb{E}[X]$ is a constant. Therefore, $g(\mu)$ is a constant, and $\nu(\mu)$ is also a constant (the slope of the supporting line at the point μ). Expectation of a constant is the constant itself, and constants can be factored out of expectations:

$$\mathbb{E}[g(X)] \geq g(\mu) + \nu(\mu)\mathbb{E}[X - \mu]$$

Finally, we evaluate the remaining expectation:

$$\mathbb{E}[X - \mu] = \mathbb{E}[X] - \mathbb{E}[\mu] = \mu - \mu = 0$$

Substituting this back into the inequality:

$$\mathbb{E}[g(X)] \geq g(\mu) + \nu(\mu) \cdot 0$$

$$\mathbb{E}[g(X)] \geq g(\mu)$$

Replacing μ with $\mathbb{E}[X]$ gives the celebrated result:

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

□

2.3 Lyapunov's Inequality: Comparing Moments

Lyapunov's inequality relates the different "absolute moments" or L_p norms of a random variable. It formalizes the idea that if a higher moment exists, then all lower moments must also exist, and it provides a comparison between their magnitudes.

Definition 2.5 (L_p Norm). For a random variable X and $p \geq 1$, the L_p norm of X is defined as $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$. (We assume the expectation exists).

Theorem 2.6 (Lyapunov's Inequality). *Let X be a random variable. If $\mathbb{E}[|X|^p] < \infty$ for some $p > 0$, then for any $0 < q \leq p$, we have $\mathbb{E}[|X|^q] < \infty$ and*

$$\|X\|_q \leq \|X\|_p$$

or equivalently,

$$(\mathbb{E}[|X|^q])^{1/q} \leq (\mathbb{E}[|X|^p])^{1/p}$$

Proof. Let's consider the case $q < p$. Define $r = p/q$. Since $p > q > 0$, we have $r > 1$. Consider the function $g(y) = |y|^r = y^r$ for $y \geq 0$. Since $r > 1$, this function is **convex** on $[0, \infty)$. (Its second derivative is $g''(y) = r(r-1)y^{r-2}$, which is ≥ 0 since $r > 1$). Now, let $Y = |X|^q$. Since $q > 0$, Y is a non-negative random variable, $Y \geq 0$. Apply Jensen's inequality (Theorem 2.4) to the convex function $g(y) = y^r$ and the non-negative random variable $Y = |X|^q$:

$$\mathbb{E}[g(Y)] \geq g(\mathbb{E}[Y])$$

Substituting $g(y) = y^r$ and $Y = |X|^q$:

$$\mathbb{E}[(|X|^q)^r] \geq (\mathbb{E}[|X|^q])^r$$

Replace r with p/q:

$$\mathbb{E}[|X|^{q \cdot (p/q)}] \geq (\mathbb{E}[|X|^q])^{p/q}$$

$$\mathbb{E}[|X|^p] \geq (\mathbb{E}[|X|^q])^{p/q}$$

Since both sides are non-negative, we can raise both sides to the positive power $1/p$. Note that $(p/q) \cdot (1/p) = 1/q$.

$$(\mathbb{E}[|X|^p])^{1/p} \geq \left((\mathbb{E}[|X|^q])^{p/q} \right)^{1/p}$$

$$(\mathbb{E}[|X|^p])^{1/p} \geq (\mathbb{E}[|X|^q])^{1/q}$$

Rearranging gives the standard form:

$$\|X\|_q \leq \|X\|_p$$

The case $q = p$ is trivial. The proof also implicitly shows that if $\mathbb{E}[|X|^p]$ is finite, $\mathbb{E}[|X|^q]$ must also be finite for $q < p$, otherwise the inequality could not hold. $\square$

Corollary 2.7 (Existence of Lower Moments). *If the p-th absolute moment of X exists (i.e., $\mathbb{E}[|X|^p] < \infty$) for some $p > 0$, then the q-th absolute moment $\mathbb{E}[|X|^q]$ also exists and is finite for all $0 < q \leq p$.*

3 Generating Random Samples

A fundamental task in computational statistics and simulation is generating random numbers that follow a specific probability distribution. While software libraries provide generators for common distributions (Uniform, Normal, Exponential, etc.), we often need methods to sample from arbitrary or less common distributions. Many techniques rely on the ability to generate random numbers from the standard uniform distribution, $U(0,1)$.

3.1 The Inverse Transform Method

The Inverse Transform Method is a general technique to convert a $U(0,1)$ random variable into a random variable with a desired CDF, F. It relies on using the inverse of the CDF (also known as the quantile function).

3.1.1 Sampling from Discrete Distributions

For a discrete random variable X taking values $x_1, x_2, \ldots$ with CDF F, we can generate a sample as follows:

Lemma 3.1 (Inverse Transform - Discrete Case). *Let X be a discrete random variable taking values $\{x_j\}_{j \geq 1}$ with CDF $F(x)$. Define $F(x_0) = 0$. Let $V \sim U(0,1)$. Define a random variable Y by setting $Y = x_j$ if $F(x_{j-1}) \leq V < F(x_j)$. Then Y has the same distribution as X. Equivalently, $Y = \sum_{j \geq 1} x_j \mathbf{1}_{[F(x_{j-1}), F(x_j))}(V)$.*

Proof. The intervals $[F(x_{j-1}), F(x_j))$ partition the interval $[0,1)$ because the CDF F increases in steps at the values x_j. For any value $v \in [0,1)$ drawn from V, it must fall into exactly one of these intervals. The probability that the generated variable Y takes the value x_j is:

$$\mathbb{P}(Y = x_j) = \mathbb{P}(F(x_{j-1}) \leq V < F(x_j))$$

Since $V \sim U(0,1)$, the probability of V falling into an interval $[a, b) \subseteq [0,1)$ is simply the length of the interval, $b - a$.

$$\mathbb{P}(Y = x_j) = F(x_j) - F(x_{j-1})$$

This is exactly the probability mass $p(x_j) = \mathbb{P}(X = x_j)$ for the original variable X. Thus, Y follows the same distribution as X. □

Example 3.2 (Lecture Example: Sampling Discrete Uniform). Suppose we want to sample X which takes values $1, 2, 3$ with $P(X = 1) = P(X = 2) = P(X = 3) = 1/3$. The CDF is $F(1) = 1/3, F(2) = 2/3, F(3) = 1$. Let $x_1 = 1, x_2 = 2, x_3 = 3$. Define $F(x_0) = 0$. The intervals are $[F(x_0), F(x_1)) = [0, 1/3), [F(x_1), F(x_2)) = [1/3, 2/3)$, and $[F(x_2), F(x_3)) = [2/3, 1)$.

Algorithm: 1. Generate $V \sim U(0,1)$. 2. If $0 \leq V < 1/3$, set $Y = x_1 = 1$. 3. If $1/3 \leq V < 2/3$, set $Y = x_2 = 2$. 4. If $2/3 \leq V < 1$, set $Y = x_3 = 3$.

Example Realization: Suppose we generate $V = 0.5$. Since $1/3 \approx 0.333$ and $2/3 \approx 0.667$, we check: Is $0 \leq 0.5 < 1/3$? No. Is $1/3 \leq 0.5 < 2/3$? Yes. Therefore, we set our sample $Y = 2$.

3.1.2 Sampling from Continuous Distributions

The method is particularly elegant for continuous distributions with strictly increasing (and thus invertible) CDFs.

Lemma 3.3 (Inverse Transform - Continuous Case). *Let X be a continuous random variable with strictly increasing CDF $F(x)$. Let $F^{-1}(v)$ be the inverse CDF (quantile function). If $V \sim U(0,1)$, then the random variable $Y = F^{-1}(V)$ has CDF $F(x)$, i.e., Y has the same distribution as X.*

Proof. Let $F_Y(y)$ be the CDF of Y. For any y in the support of X:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F^{-1}(V) \leq y)$$

Since F is strictly increasing, we can apply F to both sides without changing the inequality:

$$F_Y(y) = \mathbb{P}(F(F^{-1}(V)) \leq F(y)) = \mathbb{P}(V \leq F(y))$$

Because $V \sim U(0, 1)$, its CDF is $F_V(v) = v$ for $v \in [0, 1]$. Since $F(y)$ is a value between 0 and 1 (it's a CDF value), $\mathbb{P}(V \leq F(y))$ is simply $F(y)$.

$$F_Y(y) = F(y)$$

Thus, Y has the same CDF as X. $\square$

Lemma 3.4 (Probability Integral Transform). *Conversely, if X is a continuous random variable with strictly increasing CDF $F(x)$, then the random variable $V = F(X)$ follows the $U(0, 1)$ distribution.*

Proof. Let $F_V(v)$ be the CDF of $V = F(X)$. For $v \in [0, 1]$:

$$F_V(v) = \mathbb{P}(V \leq v) = \mathbb{P}(F(X) \leq v)$$

Since F is strictly increasing, its inverse F^{-1} exists. Applying F^{-1} to both sides:

$$F_V(v) = \mathbb{P}(F^{-1}(F(X)) \leq F^{-1}(v)) = \mathbb{P}(X \leq F^{-1}(v))$$

By the definition of the CDF F for variable X, $\mathbb{P}(X \leq x') = F(x')$. Here, $x' = F^{-1}(v)$.

$$F_V(v) = F(F^{-1}(v)) = v$$

Since $F_V(v) = v$ for $v \in [0, 1]$, V has the CDF of a $U(0, 1)$ random variable. $\square$

Example 3.5 (Lecture Example: Sampling Exponential Distribution). We want to generate a sample from the Exponential distribution with rate parameter $\lambda > 0$. The CDF is $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$. This CDF is continuous and strictly increasing on $(0, \infty)$. We need to find the inverse CDF, $F^{-1}(v)$. Set $v = F(x)$:

$$
\begin{aligned}
v &= 1 - e^{-\lambda x} \\
e^{-\lambda x} &= 1 - v \\
-\lambda x &= \ln(1 - v) \quad \text{(for } v \in [0, 1)) \\
x &= -\frac{1}{\lambda} \ln(1 - v)
\end{aligned}
$$

So, $F^{-1}(v) = -\frac{1}{\lambda} \ln(1 - v)$.

Simulation Procedure: 1. Generate a random number V from $U(0, 1)$. 2. Calculate $X = F^{-1}(V) = -\frac{1}{\lambda} \ln(1 - V)$. 3. The resulting value X is a random sample from the Exponential(λ) distribution.

Alternative Form: Since $\ln(1 - v) = -\ln(\frac{1}{1-v})$, we can also write $F^{-1}(v) = \frac{1}{\lambda} \ln(\frac{1}{1-v})$. Furthermore, if $V \sim U(0, 1)$, then $V' = 1 - V$ is also distributed $U(0, 1)$. So, a computationally common way to generate the sample is $X = -\frac{1}{\lambda} \ln(V')$, where $V' \sim U(0, 1)$.

Remark 3.6. The Inverse Transform Method provides a universal way to sample from any distribution whose inverse CDF can be computed. Its practical limitation lies in finding and evaluating F^{-1}.

— End of Lecture Material —

Next week, we will transition to studying multiple random variables simultaneously, introducing the concept of Random Vectors and Joint Distributions.