Chebyshev's Theorem

Zane Billings

Department of Mathematics and Computer Science,
Western Carolina University

12 September, 2019

Overview

Table of Contents

Chebyshev's Theorem

Theorem

For any random variable X with mean μ and finite non-zero variance σ^2,

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

Chebyshev's Theorem

Theorem

For any random variable X with mean μ and finite non-zero variance σ^2,

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

- The Empirical Rule (or 68-95-99 Rule) provides good bounds for the probability that the value of a normally distributed variable lies within a given range of the mean.

Theorem

For any random variable X with mean μ and finite non-zero variance σ^2,

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

- The Empirical Rule (or 68-95-99 Rule) provides good bounds for the probability that the value of a normally distributed variable lies within a given range of the mean.
- Chebyshev's Theorem generalizes this bound to nearly all random variables **regardless of distribution**.

Table of Contents

Random Variables

Definition

A **random variable** is a function from a sample space, S, to $\mathbb{R}$. [3]

Random Variables

Definition

A **random variable** is a function from a sample space, S, to $\mathbb{R}$. [3]

Definition

A random variable is said to be **discrete** if it can only take on a finite or countably infinite set of possible values, or **continuous** if it can take on any value in a specified interval.

Random Variables

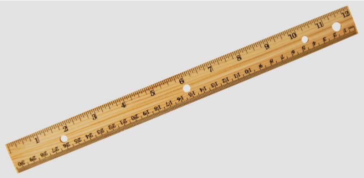
Definition

A **random variable** is a function from a sample space, S, to $\mathbb{R}$. [3]

Definition

A random variable is said to be **discrete** if it can only take on a finite or countably infinite set of possible values, or **continuous** if it can take on any value in a specified interval.

(a) Discrete



(b) Continuous

Figure: Random variable examples.

Distribution Functions of a Random Variable

Definition

Every random variable has a **cumulative distribution function** (or cdf) defined by

$$F_X(x) = P_X(X \leq x), \forall x. \text{ [1]}$$

$F(x)$ must satisfy three properties to be a cdf:

1. $\lim_{x \to -\infty} F(x) = 0; \; \lim_{x \to \infty} F(x) = 1.$
2. $F(x)$ is a nondecreasing function.
3. $F(x)$ is right-continuous, i.e. $\lim_{x \to x_0^+} F(x) = F(x_0)$ for all $x_0 \in \mathbb{R}$.

Distribution Functions of a Random Variable

Definition

Every random variable has a **cumulative distribution function** (or cdf) defined by

$$F_X(x) = P_X(X \leq x), \forall x. \text{ [1]}$$

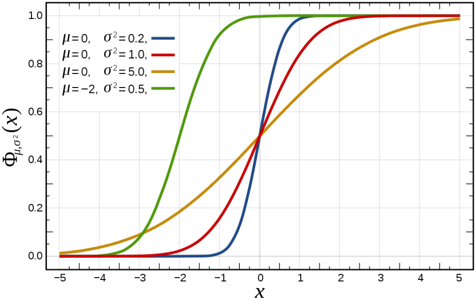


Figure: The CDF of the normal distribution with varying parameters.

Distribution Functions of a Random Variable

Definition

The **probability density function** (pdf) of a continuous random variable X is the function f for which

$$f_X(x) = \frac{d}{dx} F_X(x). \text{ [1]}$$

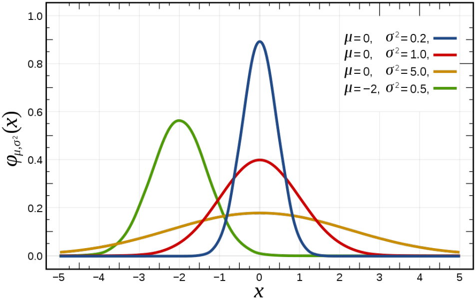


Figure: The PDF of the normal distribution with varying parameters.

Expected Value and Variance of a Random Variable

Definition

The **expected value** or expectation of a random variable X (also called the mean) is defined as:

$$\mathbb{E}[X] = \mu = \int_{\mathcal{S}} x f_X(x)\ dx$$

where $f_X(x)$ is the density function associated with X. [2]

Expected Value and Variance of a Random Variable

Definition

The **expected value** or expectation of a random variable X (also called the mean) is defined as:

$$\mathbb{E}[X] = \mu = \int_{\mathcal{S}} x f_X(x)\ dx$$

where $f_X(x)$ is the density function associated with X. [2]

Definition

The **variance** of a random variable X is defined as:

$$\mathbb{E}[(X-\mu)^2] = \sigma^2 = \int_{\mathcal{S}} (x-\mu)^2 f_X(x)\ dx$$

where $f_X(x)$ is the density function associated with X. [2]

Two properties we will need

Expected value is linear.

A linear operation "plays nicely" with addition and scalar multiplication:

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

A formal proof follows from the fact that integration is linear [2].

Two properties we will need

Expected value is linear.

A linear operation "plays nicely" with addition and scalar multiplication:

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

A formal proof follows from the fact that integration is linear [2].

Support of a random variable.

Suppose the pdf of the random variable X is strictly positive on the set $\mathcal{S}$ and zero elsewhere. The set $\mathcal{S}$ is called the **support set** (or support) of the distribution. Importantly:

$$\int_{\mathbb{R}} x f_X(x)\ dx = \int_{\mathcal{S}} x f_X(x)\ dx. \text{ [1]}$$

Table of Contents

Lemma 1.

Lemma

$$\Pr(X \geq a) = 1 \implies \mathbb{E}[X] \geq a.$$

Lemma 1.

Lemma

$$\Pr(X \geq a) = 1 \implies \mathbb{E}[X] \geq a.$$

Proof.

- Let X be a continuous random variable s.t. $P(X \geq a) = 1$, and suppose that $\mathcal{S}$ is the support of X.

Lemma

$$\Pr(X \geq a) = 1 \implies \mathbb{E}[X] \geq a.$$

Proof.

- Let X be a continuous random variable s.t. $P(X \geq a) = 1$, and suppose that $\mathcal{S}$ is the support of X.
- Since $\Pr(X \geq a) = 1$, we have $x \in \mathcal{S} \geq a$ for all x: if we assume we have some $x < a \in \mathcal{S}$, $\Pr(X \leq x) \subset \Pr(X < a) = \Pr((X \geq a)^C) = 0$, so $\Pr(X \leq x) = 0$, and $x \notin \mathcal{S}$.

Lemma 1.

Lemma

$$\Pr(X \geq a) = 1 \implies \mathbb{E}[X] \geq a.$$

Proof.

- Let X be a continuous random variable s.t. $P(X \geq a) = 1$, and suppose that $\mathcal{S}$ is the support of X.
- Since $\Pr(X \geq a) = 1$, we have $x \in \mathcal{S} \geq a$ for all x: if we assume we have some $x < a \in \mathcal{S}$, $\Pr(X \leq x) \subset \Pr(X < a) = \Pr((X \geq a)^C) = 0$, so $\Pr(X \leq x) = 0$, and $x \notin \mathcal{S}$.
- Therefore,

$$\mathbb{E}[X] = \int_{\mathcal{S}} x f_X(x) \ dx \geq \int_{\mathcal{S}} a f_X(x) \ dx = a \int_{\mathcal{S}} f_X(x) \ dx = a.$$

□

Lemma 2

Lemma.

$$\Pr(X \geq Y) = 1 \implies \mathbb{E}[X] \geq \mathbb{E}[Y].$$

Lemma 2

Lemma.

$$\Pr(X \geq Y) = 1 \implies \mathbb{E}[X] \geq \mathbb{E}[Y].$$

Proof.

- Let X and Y be random variables such that $\Pr(X \geq Y) = 1$, and let Z be a random variable such that $Z = X - Y$.

Lemma 2

Lemma.

$$\Pr(X \geq Y) = 1 \implies \mathbb{E}[X] \geq \mathbb{E}[Y].$$

Proof.

- Let X and Y be random variables such that $\Pr(X \geq Y) = 1$, and let Z be a random variable such that $Z = X - Y$.
- Since $\Pr(X \geq Y) = 1$, we have $\Pr(X - Y \geq 0) = 1$, and thus $\Pr(Z \geq 0) = 1$. By Lemma 1, we have that $\mathbb{E}[Z] \geq 0$.

Lemma 2

Lemma.

$$\Pr(X \geq Y) = 1 \implies \mathbb{E}[X] \geq \mathbb{E}[Y].$$

Proof.

- Let X and Y be random variables such that $\Pr(X \geq Y) = 1$, and let Z be a random variable such that $Z = X - Y$.
- Since $\Pr(X \geq Y) = 1$, we have $\Pr(X - Y \geq 0) = 1$, and thus $\Pr(Z \geq 0) = 1$. By Lemma 1, we have that $\mathbb{E}[Z] \geq 0$.
- $\mathbb{E}[Z] \geq 0 \implies \mathbb{E}[X - Y] \geq 0 \implies \mathbb{E}[X] - \mathbb{E}[Y] = 0$, by the linearity of expected value.

Lemma 2

Lemma.

$$\Pr(X \geq Y) = 1 \implies \mathbb{E}[X] \geq \mathbb{E}[Y].$$

Proof.

- Let X and Y be random variables such that $\Pr(X \geq Y) = 1$, and let Z be a random variable such that $Z = X - Y$.
- Since $\Pr(X \geq Y) = 1$, we have $\Pr(X - Y \geq 0) = 1$, and thus $\Pr(Z \geq 0) = 1$. By Lemma 1, we have that $\mathbb{E}[Z] \geq 0$.
- $\mathbb{E}[Z] \geq 0 \implies \mathbb{E}[X - Y] \geq 0 \implies \mathbb{E}[X] - \mathbb{E}[Y] = 0$, by the linearity of expected value.
- Therefore, $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

$\square$

Table of Contents

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof.

- Suppose X is a non-negative random variable and $a \in \mathbb{R}_{>0}$.

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof.

- Suppose X is a non-negative random variable and $a \in \mathbb{R}_{>0}$.
- Let I_a be the indicator variable for the event $X \geq a$. That is,

$$I_a = \begin{cases} 1, & \text{if } X \geq a; \\ 0, & \text{if } X < a. \end{cases}$$

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof.

- Suppose X is a non-negative random variable and $a \in \mathbb{R}_{>0}$.
- Let I_a be the indicator variable for the event $X \geq a$. That is,

$$I_a = \begin{cases} 1, & \text{if } X \geq a; \\ 0, & \text{if } X < a. \end{cases}$$

- Then, we see (since I_a is discrete by definition):

$$\mathbb{E}[I_a] = (1)\Pr(X \geq a) + (0)\Pr(X < a) = \Pr(X \geq a).$$

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

- Now, suppose $X \geq a$. Then, $I_a = 1$ and $aI_a = a$. Hence, $X \geq aI_a$.

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

- Now, suppose $X \geq a$. Then, $I_a = 1$ and $aI_a = a$. Hence, $X \geq aI_a$.
- However, if $X < a$, $I_a = 0$, and since X is non-negative, $X \geq aI_a$.

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

- Now, suppose $X \geq a$. Then, $I_a = 1$ and $aI_a = a$. Hence, $X \geq aI_a$.
- However, if $X < a$, $I_a = 0$, and since X is non-negative, $X \geq aI_a$.
- So, we see that X is always $\geq aI_a$, that is, $\Pr(X \geq aI_a) = 1$. Thus, by Lemma 2, $\mathbb{E}[X] \geq \mathbb{E}[aI_a]$.

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

- Now, suppose $X \geq a$. Then, $I_a = 1$ and $aI_a = a$. Hence, $X \geq aI_a$.
- However, if $X < a$, $I_a = 0$, and since X is non-negative, $X \geq aI_a$.
- So, we see that X is always $\geq aI_a$, that is, $\Pr(X \geq aI_a) = 1$. Thus, by Lemma 2, $\mathbb{E}[X] \geq \mathbb{E}[aI_a]$.
- By the linearity of expected value, we have $\mathbb{E}[X] \geq a\mathbb{E}[I_a]$.

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

Now, we have:

$$\mathbb{E}[X] \geq a \cdot \mathbb{E}[I_a]$$

$\square$

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

Now, we have:

$$\mathbb{E}[X] \geq a \cdot \mathbb{E}[I_a]$$
$$\frac{\mathbb{E}[X]}{a} \geq \mathbb{E}[I_a]$$

$\square$

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

Now, we have:

$$\begin{aligned} \mathbb{E}[X] &\geq a \cdot \mathbb{E}[I_a] \\ \frac{\mathbb{E}[X]}{a} &\geq \mathbb{E}[I_a] \\ \frac{\mathbb{E}[X]}{a} &\geq \Pr(X \geq a) \end{aligned}$$

$\square$

Markov's Inequality

Theorem

For any non-negative random variable X, and $a \in \mathbb{R}_{>0}$,

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

Proof (Cont.)

Now, we have:

$$\mathbb{E}[X] \geq a \cdot \mathbb{E}[I_a]$$
$$\frac{\mathbb{E}[X]}{a} \geq \mathbb{E}[I_a]$$
$$\frac{\mathbb{E}[X]}{a} \geq \Pr(X \geq a)$$

(This is exactly what we wanted to show!) $\square$

Table of Contents

Chebyshev's Theorem

For any random variable X with mean μ and finite non-zero variance σ^2,

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

Proof.

Let Y be a random variable such that $Y = (X - \mu)^2$, where X is a random variable with mean μ and finite non-zero variance, σ^2. Finally, let $k \in \mathbb{R}_{>0}$.

$\square$

Chebyshev's Theorem

Chebyshev's Theorem

For any random variable X with mean μ and finite non-zero variance σ^2,

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

Proof.

Let Y be a random variable such that $Y = (X - \mu)^2$, where X is a random variable with mean μ and finite non-zero variance, σ^2. Finally, let $k \in \mathbb{R}_{>0}$.

$$\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \mathrm{Var}[X]$$

$\square$

Chebyshev's Theorem

Chebyshev's Theorem

For any random variable X with mean μ and finite non-zero variance σ^2,

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

Proof.

Let Y be a random variable such that $Y = (X - \mu)^2$, where X is a random variable with mean μ and finite non-zero variance, σ^2. Finally, let $k \in \mathbb{R}_{>0}$.

$$\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \mathrm{Var}[X]$$

$$\Pr((X - \mu)^2 \geq k^2\sigma^2) \leq \frac{\mathbb{E}[Y]}{k\sigma^2} = \frac{\mathrm{Var}[X]}{k\sigma^2} = \frac{\sigma^2}{k^2\sigma^2}$$

$\square$

Chebyshev's Theorem

Chebyshev's Theorem

For any random variable X with mean μ and finite non-zero variance σ^2,

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

Proof.

Let Y be a random variable such that $Y = (X - \mu)^2$, where X is a random variable with mean μ and finite non-zero variance, σ^2. Finally, let $k \in \mathbb{R}_{>0}$.

$$\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \mathrm{Var}[X]$$

$$\Pr((X - \mu)^2 \geq k^2\sigma^2) \leq \frac{\mathbb{E}[Y]}{k\sigma^2} = \frac{\mathrm{Var}[X]}{k\sigma^2} = \frac{\sigma^2}{k^2\sigma^2}$$

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

$\square$

References

George Casella and Roger L. Berger.
Statistical Inference.
Duxbury, second edition, 2002.

Sheldon M Ross.
Introduction to probability models.
Academic Press, Inc., fourth edition, 1989.

Dennis D. Wackerly, William Medenhall III, and Richard L. Scheaffer.
Mathematical Statistics with Applications.
Thomson, seventh edition, 2008.

Image Credits

- Coins from https://www.futilitycloset.com/2015/05/02/three-of-a-kind/
- Ruler from https://www.inchcalculator.com/ruler-information-uses-types/
- Normal CDF and PDF are public domain images from WikiCommons.