Probability 1 - Lecture Notes

Philip Warton

November 4, 2020

1 Markov Inequality

Suppose there is a distribution for which we don't know the probability mass function, and we do not know the variance, but we do know it's expectation, $E[x]$. What can we say about that probability? Can we bound it?

Theorem 1.1 (Markov Inequality)**.** *If X is a random variable that takes only non-negative values, then for any $\alpha > 0$,*

$$P(X \leq \alpha) \leqslant \frac{E[x]}{\alpha}$$

Proof.

$$P(X \geq \alpha) = \sum_{k:k \geq \alpha} p(\alpha) \leq \sum_{k:k \geq \alpha} \frac{k}{\alpha} p(k) = \frac{1}{\alpha} \sum_{k:k \geq \alpha} k \cdot p(k) \leq \frac{1}{\alpha} \sum_{k:k \geq 0} k \cdot p(k) = \frac{E[X]}{\alpha}$$

□

Note that this would likely work under integration for a continuous random variable.

Theorem 1.2 (Chebyshev Inequality)**.** *If X is a random variable with a finite mean μ and variance, then for any $\kappa > 0$,*

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

2 Continuous Random Variables

Definition 2.1. We say that X is a continuous random variable if there exists a nonnegative function $f(x)$ defined for all real x such that for any $a \leq b$

$$P(a \leq X \leq b) = \in_a^b f(x)dx$$

Such a function $f(x)$ is the probability density function of X Figure 1.

First notice that the prboability density function must be non-negative, because it is impossible to have a negative probability by definition axiomatically. There are some properties of these functions that we wish enumerate now:

(i) $\displaystyle\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X < \infty) = 1$

(ii) $\displaystyle P(X = a) = \int_a^a f(x)dx = 0 \forall a \in \mathbb{R}$

(iii) $\displaystyle P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx$

We can restate this definition by saying, $f(x)$ is a probability density function $\Leftrightarrow$ $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$. Even though $P(X = a) = 0$ for every real number a, since the real numbers are uncountable, we do not violate any of our axioms of probability. Since $P(S) = 1$ for any sample space S, it follows that $P(-\inf \leq X \leq \inf) = 1$.

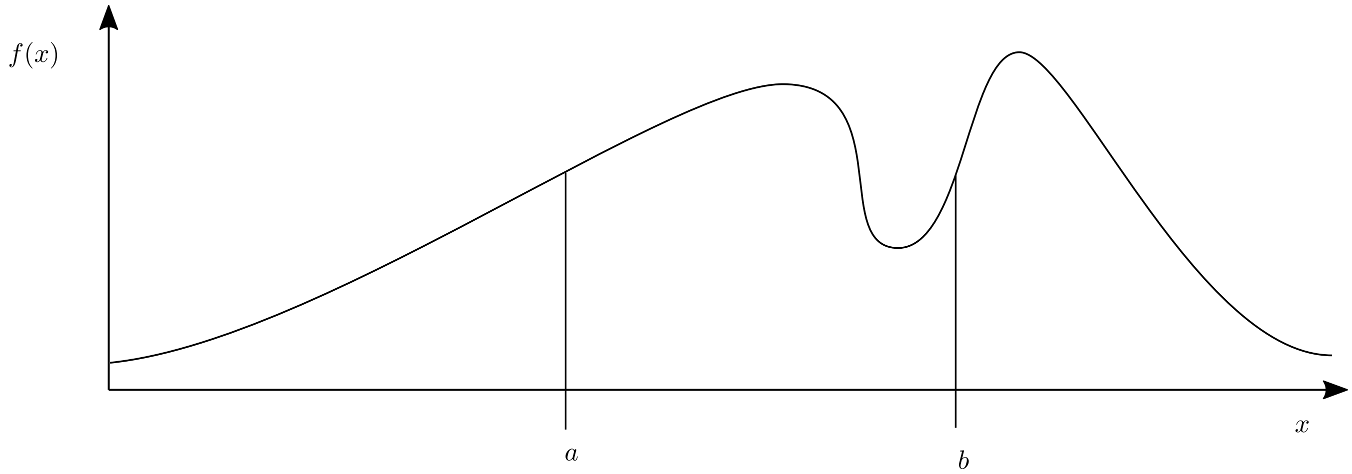

Figure 1: Probability Density Funciton

Let us take the example of the following function:

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We know that this function will integrate to 1 over $\mathbb{R}$. Scaling, this function by λ we get another probability density function.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We obtain the same exact area, so we still have a valid probability density function so long as $\lambda > 0$. This is called an exponential random variable. It is a continuous analogue to the geometric random variable in the discrete case. Then it also carries the property of memorylessness, which means that $P(X > a + b | X > a) = P(X > b)$.